AD-A114 550 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1

CHOICE OF RESPONSE SURFACE DESIGN AND ALPHABETIC OPTIMALITY.(U)

FEB 82 G E BOX DAAG29-80-C-0041

UNCLASSIFIED MRC-TSR-2333 NL

1 OF 1

AD A114550

END

DATE FILMED

6 82

DTIC

MRC Technical Summary Report # 2333

CHOICE OF RESPONSE SURFACE DESIGN AND ALPHABETIC OPTIMALITY

George E. P. Box

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

February 1982

(Received September 10, 1981)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

CHOICE OF RESPONSE SURFACE DESIGN AND ALPHABETIC OPTIMALITY

George E. P. Box

Technical Summary Report # 2333

February 1982

ABSTRACT

It is argued that the specification of problems of experimental design (and in particular, of response surface design) should depend on scientific context. The specification for a widely developed theory of "alphabetic optimality" for response surface applications is analyzed and found to be unduly limiting. Ways in which designs might be chosen to satisfy a set of criteria of greater scientific relevance are suggested. Detailed consideration is given to regions of operability and interest, to the design information function, to sensitivity of criteria to size and shape of the region, and to the effect of bias. Problems are discussed of checking for lack of fit, sequential assembly, orthogonal blocking, estimation of error, estimation of transformations, robustness to bad values, using minimum numbers of points, and employing simple data patterns.

AMS (MOS) Subject Classification: 62K05

Key Words: D-optimality, G-optimality, A-optimality, E-optimality, response surface designs, information function, bias, sequential assembly, robustness.

Work Unit Number 4 - Statistics and Probability

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

Response surface methods for investigating empirical relationships, and in particular for finding the experimental conditions for which some response is maximized, were introduced some thirty years ago. Since that time these methods have been widely applied, and much research has gone into improving them, in particular the selection of suitable experimental designs. One prolific line of mathematical research has concerned particular optimality criteria (e.g. D, G, A, and E optimality) called here "alphabetic optimality." The assumptions and specifications which motivate alphabetic optimality are discussed, and found to be unduly limiting so far as response surface design is concerned. An approach having greater scientific relevance is discussed.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

Choice of Response Surface Design and Alphabetic Optimality

George E. P. Box

University of Wisconsin - Madison

1. INTRODUCTION

There seems no doubt that of all the activities in which the statistician can engage, that of designing experiments is by far the most important, since it is here that the actual mode of generation of scientific data is decided.

The importance of practice in guiding the development of the theory of experimental design [45] is clearly seen from the time of its invention. Fisher was engaged by Russell [16] on a temporary basis at Rothamsted Experimental Station in 1919 "to examine our data and elicit further information that we had missed." Records were available from the ongoing Broadbalk experiment in which particular combinations of fertilizers had been consistently applied to 13 plots for a period of almost 70 years. In his analysis ([22], [24]), Fisher attempted to relate yield to fertilizer combination, to weather, and in particular, to rainfall. The method he used was multiple regression with distributed lag models, involving an ingenious employment of orthogonal polynomials which led to important advances in the theory of regression analysis, and in particular its distribution theory.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

With only the crudest of computational aids, the work must have been burdensome, making it all the more frustrating to discover that, however ingenious the analysis, the inherent nature of the data ensured that the answers to many questions were inaccessible. A comprehension of the logical problems in drawing conclusions from such analyses led naturally to speculation on how some of the difficulties might be overcome by appropriate design. These ideas were further stimulated by the Analysis of Variance, which Fisher introduced in 1923 with W.A. Mackenzie [23] for the elucidation of what was clearly a most unsatisfactory design which he had had no part in choosing. Thereafter, as Fisher gradually acquired more influence in the setting up of field trials, the principles of replication, randomization and their application to randomized blocks, latin squares and factorial designs quickly evolved out of the actual planning, running, and analysis of a series of experimental designs of increasing complexity and beauty.

The practical context of scientific experimentation continued to produce important theoretical advances when Yates came to Rothamsted in 1931, leading in particular to important developments in the design and analysis of complex factorial designs and their associated systems of confounding ([44], [46]) and to the introduction of incomplete block designs.

My own experience with experimental design began during the Second World War. I worked at the Chemical Defense Experimental Station in England with a group of medical research workers who were attempting, using animals and volunteers, to find ways to combat the effects of poison gas and other toxic agents. At this time it was believed that these agents might be used not only against the military, but also against the civilian population. It was important therefore that our work should progress as rapidly as possible. I found myself a part of evolving investigations which employed

sequences of experiments which I designed and whose nature needed to adapt to changing needs at different stages of the study. The designs employed were randomized blocks, balanced incomplete blocks, latin squares, and factorials. Later, during my eight years as a statistician with Imperial Chemical Industries, my role was again as a member of various scientific teams tackling evolving problems with sequences of designs. Many of the problems were similar to those I had previously encountered, and again employed the (by now) standard designs of Fisher and Yates. However, some investigations directly concerned with the improvement of chemical processes at the lab, pilot plant, and full scale, seemed to require additional methods, which however, still drew on the fundamental principles laid down by the originators of experimental design. This led to the development of what has come to be called response surface methodology. See for example [4], [14], [15], [30], [31], and [39].

Suppose some response η of interest is believed to be locally approximated by a polynomial of low degree in k continuous experimental variables $\underset{\sim}{x} = (x_1, x_2, \ldots, x_k)'$. To fit such a function we need appropriate experimental designs. Let us call a design suitable for estimating a general polynomial of degree d a d̲th order design in k variables. Thus a design suitable for fitting the function

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

would be a second order design in two variables.

One route for choosing such designs, which has generated an enormous amount of mathematical research over the last twenty or so years, we shall refer to as the "alphabetic optimality" approach. For reasons I will explain,

I have reservations about the usefulness of this approach so far as response surface designs are concerned. For completeness, a brief summary of some of the main ideas are set out below ([32], [33], [34], [35], [36], [42], [43]).

2. SOME ASPECTS OF OPTIMAL DESIGN THEORY FOR CONTINUOUS EXPERIMENTAL VARIABLES

Consider a response η which is supposed to be an exactly known function $\eta = \underset{\sim}{x}'\underset{\sim}{\beta}$ linear in p coefficients $\underset{\sim}{\beta}$, where $\underset{\sim}{x} = \{f_1(\underset{\sim}{\chi}), f_2(\underset{\sim}{\chi}), \ldots, f_p(\underset{\sim}{\chi})\}'$ is a vector of p functions of k experimental variables $\underset{\sim}{\chi}$. Suppose a design is to be run defining n sets of k experimental conditions given by the n x k design matrix $\{\underset{\sim}{\chi}_u\}$ and yielding n observations $\{y_u\}$, so that

$$\eta_u = \underset{\sim}{x}_u'\underset{\sim}{\beta} \qquad (u = 1, 2, \ldots, n)$$

where $y_u - \eta_u = \varepsilon_u$ is distributed $N(0,\sigma^2)$ and the n x p matrix $\underset{\sim}{X} = \{\underset{\sim}{x}_u'\}$.

The elements of $\{c_{ij}\} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ are proportional to the variances and covariances of the least squares estimates $\hat{\underset{\sim}{\beta}}$. Within this specification, the problem of experimental design is that of choosing the design $\{\underset{\sim}{\chi}_u\}$ so that the elements c_{ij} are to our liking. Because there are $\frac{1}{2}p(p+1)$ of these, simplification is desirable.

A motivation for simplification is provided by considering the confidence region[1] for $\underset{\sim}{\beta}$

---

[1] Obviously there are also parallel fiducial and Bayesian rationalizations.

$$(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}})'\underset{\sim}{X}'\underset{\sim}{X}(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}) = \text{constant}$$

defining an ellipsoid in p parameters. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ are proportional to the squared lengths of the p principal axes of this ellipsoid. Suppose their maximum, arithmetic mean, and geometric mean are indicated by λ_{max}, $\bar{\lambda}$, and $\tilde{\lambda}$. Then it is illuminating to consider the transformation of the $\frac{1}{2}p(p+1)$ elements c_{ij} to a corresponding number of items as follows:

(i) $D = |X'X| = \tilde{\lambda}^{-p}$ (so that $D^{-1/2} = \tilde{\lambda}^{p/2}$ is proportional to the volume of the confidence ellipsoid).

(ii) $\underset{\sim}{H}$, a vector of p - 1 homogeneous functions of degree zero in the λ's, which measure the non-sphericity or state of ill-conditioning of the ellipsoid. In particular we might choose, for two of these, $H_1 = \bar{\lambda}/\tilde{\lambda}$ and $H_2 = \lambda_{max}/\tilde{\lambda}$, both of which would take the value unity for a spherical region.

(iii) $\frac{1}{2}p(p-1)$ independent direction cosines which determine the orientation of the orthogonal axes of the ellipsoid.

It is traditionally assumed that the $\frac{1}{2}p(p-1)$ elements concerned with orientation of the ellipsoid are of no interest, and attention has been concentrated on particular criteria which measure in some way or another the sizes of the eigenvalues, measuring some combination of size and sphericity of the confidence ellipsoid. Among these criteria are

$$D = |X'X| = \Pi\lambda_i^{-1} = \tilde{\lambda}^{-p}$$
$$A = \Sigma\lambda_i = \text{tr}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} = \sigma^{-2}\Sigma\text{var}(\hat{\beta}_i) = p\tilde{\lambda}H_1$$
$$E = \max\{\lambda_i\} = \tilde{\lambda}H_2$$

The desirability of a design, as measured by the D, A, and E criteria, increases as $\tilde{\lambda}$, $\tilde{\lambda}H_1$, and $\tilde{\lambda}H_2$ respectively, are decreased. But in practical situations, each of these criteria will take smaller and hence more desirable values as the ranges for the experimental variables χ are taken larger and larger. To cope with this problem it is usually assumed that the experimental variables $\underset{\sim}{\chi}_u$ may vary only within some exactly known region in the space of $\underset{\sim}{\chi}$, but not outside it. I will call this permissible region $\underset{\sim\sim}{RO}$.

Another characteristic of the problem which makes its study mathematically difficult is the necessary discreteness of the number of runs which can be made at any given location. In a technically brilliant paper [37], Kiefer and Wolfowitz dealt with this obstacle by introducing a continuous design measure ξ which determines the <u>proportion</u> of runs which should ideally be made at each of a number of points in the $\underset{\sim}{\chi}$ space. Realizable designs which most nearly approximated the optimal distribution could then be used in practice.

A further important result of Kiefer and Wolfowitz linked the problem of estimating $\underset{\sim}{\beta}$ with that of estimating the response η via the property of "G-optimality." G-optimal designs were defined as those which minimized the maximum value of $V(\hat{y}_x)$ <u>within</u> $\underset{\sim\sim}{RO}$. The authors were then able to show, for their measure designs, the equivalence of G- and D-optimality. Furthermore, they showed that, for such a design, within the region $\underset{\sim\sim}{RO}$, the maximum value of $n\cdot Var(\hat{y}_x)/\sigma^2$ was p, and that this value was actually attained at each of the design points.

For illustration we consider a second order measure-design in two variables; that is, a design appropriate for the fitting of the second degree polynomial of equation (1). Such a design which is both D- and G- optimal for a square region $\underset{\sim\sim}{RO}$ with vertices $(\pm1,\pm1)$ was given by Fedorov [21] (see

also Herzberg [27]). The design places 14.6% of the measure at each of the four vertices, 8.0% at each of the midpoints of the edges, and 9.6% at the origin. The design is set out in Figure 4(b).

While this approach has generated much interesting mathematics, it does not, I believe, solve the problem of choosing good response surface designs. In the hope of stimulating new initiative, I have set out below what I believe is the scientific context for response surface studies and indicated some possible lines of development.

3. THE RESPONSE SURFACE CONTEXT

As an example suppose it is desired to study some chemical system, with the object of obtaining a higher value for a response η such as yield which is initially believed to be some function $\eta = g(\underset{\sim}{\chi})$ of k continuous input variables $\underset{\sim}{\chi} = (\chi_1, \chi_2, \ldots, \chi_k)'$ such as reaction time, temperature, or concentration. As is illustrated in Figure 1, it is usually known initially that the system can be operated at some point $\underset{\sim}{\chi}_0$ in the space of $\underset{\sim}{\chi}$ and is expected to be capable of operating over some much more extensive region $\underset{\sim}{O}$ called the <u>operability region</u>, which[2] however is usually unknown or poorly known. Response surface methods are employed when the nature of the true response function $\eta = g(\underset{\sim}{\chi})$ is also unknown[3] or is inaccessible.

[2] One secondary object of the investigation may be to find out more about the operability region $\underset{\sim}{O}$.

[3] Occasionally the true functional form $\eta = g(\underset{\sim}{\chi})$ may be known, or at least conjectured, from knowledge of physical mechanisms. Typically however $g(\underset{\sim}{\chi})$ will then appear as a solution of a set of differential equations which are nonlinear in a number of parameters which may represent physical constants. Problems of nonlinear experimental design then arise which are of considerable interest although they have received comparatively little attention (see for example [13], [18], [25]).

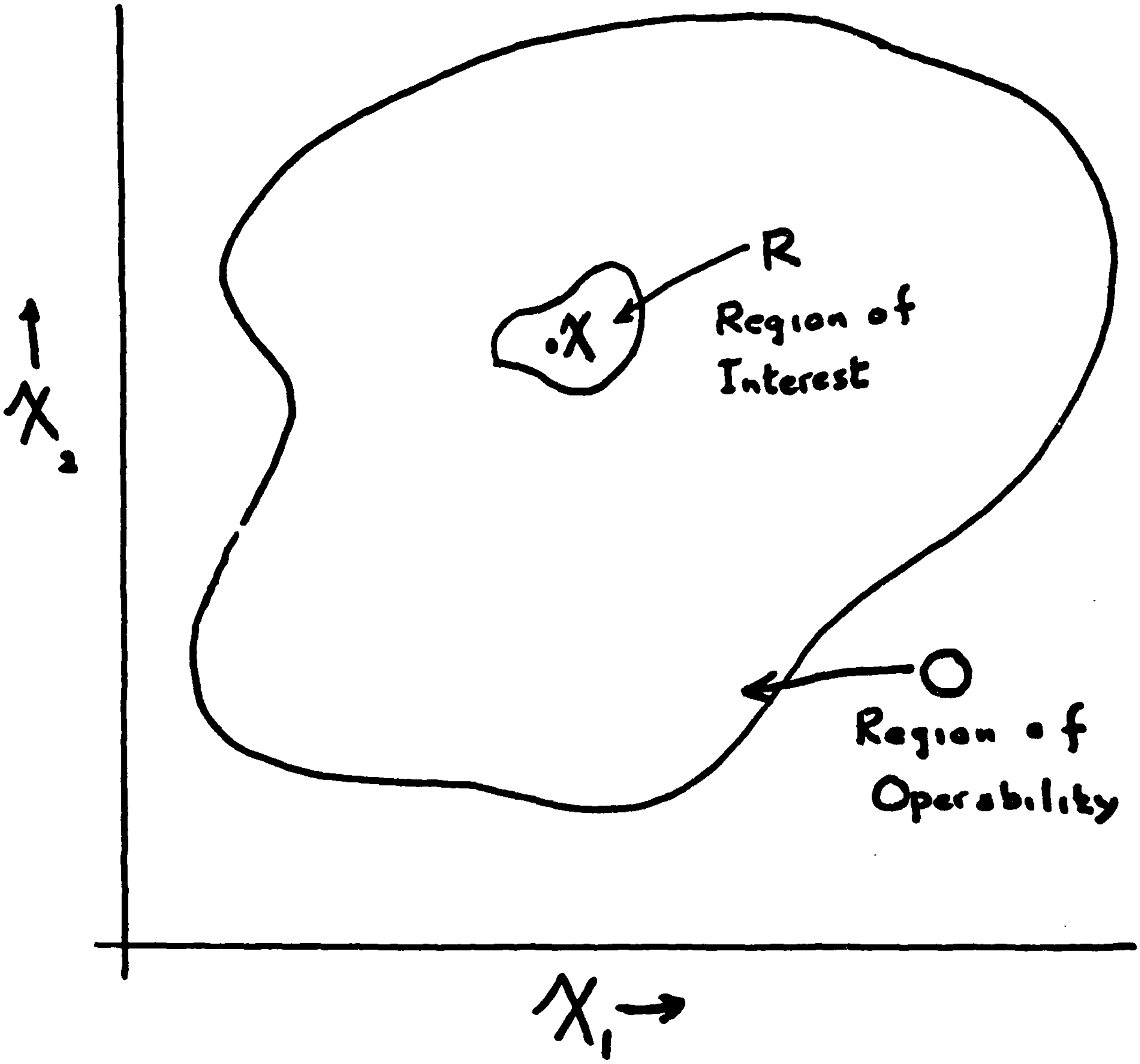


Figure 1. The current region of interest $\underset{\sim}{R}$ and the region of operability $\underset{\sim}{O}$ in the space of two continuous experimental variables X_1 and X_2.

Suppose that over some (typically much less extensive) immediate region of interest $\underset{\sim}{R}$ in the neighborhood of $\underset{\sim}{\chi}_0$ it is guessed that a "graduating" function, such as a dth degree polynomial in x,

$$\eta_{\underset{\sim}{\chi}} \doteq \underset{\sim}{x}'\underset{\sim}{\beta}$$

might provide a locally adequate approximation to the true function $\eta_{\underset{\sim}{\chi}} = g(\underset{\sim}{\chi})$ where as before $\underset{\sim}{x}$ is a p-dimensional vector of suitably transformed input variables $\underset{\sim}{x}' = \{f_1(\underset{\sim}{\chi}), f_2(\underset{\sim}{\chi}), \ldots, f_p(\underset{\sim}{\chi})\}$, and $\underset{\sim}{\beta}$ is a vector of coefficients occurring linearly that may be adjusted to approximate the unknown true response function $\eta_{\underset{\sim}{\chi}} = g(\underset{\sim}{\chi})$. Then progress may be achieved by using a sequence of such approximations. For example when a first degree polynomial approximation could be employed it might, via the method of steepest ascent, be used to find a new region of interest R_1 where, say, the yield was higher. Also a maximum in many variables is often represented by some rather complicated ridge system[4] and a second degree polynomial approximation when suitably analysed might be used to elucidate, describe, and exploit such a system.

Thus we are typically involved in using a sequence of designs, each making use of information gleaned from earlier experiments -- a characteristic typical of a much wider field of scientific investigation. This provides the

---

[4] Empirical evidence suggests this. Also, integration of sets of differential equations which describe the kinetics of chemical systems almost invariably leads to ridge systems ([4], [15], [26], [41]). See also the discussion of Figure 8.

opportunity to progressively improve not only the objective function η directly, but also the mode of gathering information about it. For example, at the ith stage, a design performed in a region $\underset{\sim}{R}_i$ may suggest that a new region $\underset{\sim}{R}_{i+1}$ is worthy of investigation (either because it can be expected to give higher values of η or because it may throw light on other important aspects of the function). But this new region may be different not only in (a) its location in the space of $\underset{\sim}{\chi}$, but (b) in its shape also (for instance because of information fed back from previous data on transformations of χ's individually or jointly), and (c) in the identity of its component space (because of feedback from the results themselves, indicating that certain variables should be dropped, and/or that new variables should be added). Thus in any realistic view of the process of investigation the dimensions, identity, location and metrics of measurement of regions of interest in the experimental space are all iteratively evolving. The problem of choosing suitable experimental designs in such a context is a difficult one. Some properties ([5], [8]) of a response surface design, any, all or some of which might in different circumstances be of importance in the above context are given in Table 1.

The design information function

Associated with requirements (1) and (2) of Table 1, consider the design variance function [11]

$$V_{\underset{\sim}{\chi}} = n\cdot\mathrm{Var}(\hat{y}_{\underset{\sim}{\chi}})/\sigma^2 = n\underset{\sim}{x}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{x}$$

or equivalently the Information Function

$$I_{\underset{\sim}{\chi}} = V_{\underset{\sim}{\chi}}^{-1} .$$

The design should:

(i) generate a satisfactory distribution of information throughout the region of interest, R;

(ii) ensure that the fitted value at x, $\hat{y}(x)$ be as close as possible to the true value at x, $\eta(x)$;

(iii) give good detectability of lack of fit;

(iv) allow transformations to be estimated;

(v) allow experiments to be performed in blocks;

(vi) allow designs of increasing order to be built up sequentially;

(vii) provide an internal estimate of error;

(viii) be insensitive to wild observations and to violation of the usual normal theory assumptions;

(ix) require a minimum number of experimental points;

(x) provide simple data patterns that allow ready visual appreciation;

(xi) ensure simplicity of calculation;

(xii) behave well when errors occur in the settings of the predictor variables, the x's;

(xiii) not require an impractically large number of predictor variable levels;

(xiv) provide a check on the 'constancy of variance' assumption.

TABLE 1. SOME ATTRIBUTES OF DESIGNS OF POTENTIAL IMPORTANCE

It is evident that if we were to make the unrealistic assumption (made in alphabetic optimality) that the graduating function $\eta = \underset{\sim}{x}'\underset{\sim}{\beta}$ is capable of <u>exactly</u> representing the true function $g(\underset{\sim}{\chi})$, then the information function would tell us all we could know about the design's ability to estimate η. For illustration, information functions and associated information contours for a 2^2 factorial used as a first order design and for a 3^2 factorial used as a second order design are shown in Figures 2 and 3, for standard variables x_1 and x_2.

4. APPLICABILITY OF ALPHABETIC OPTIMALITY

The information function for Fedorov's second order D/G-optimal design over the permissible $\underset{\sim\sim}{RO}$ region ($\pm1,\pm1$), referred to earlier, is shown in Figure 4. For illustration, this is related to the two experimental variables χ_1 = temp in ^{0}C and χ_2 = time in hours. Thus, in this particular example, $x_1 = (\chi_1 - 180)/10$, $x_2 = \chi_2 - 4$ and the $\underset{\sim\sim}{RO}$ region would permit experimentation within the limits $\chi_1 = 170 - 190$ °C and $\chi_2 = 3 - 5$ hours, but not outside these limits. In the response surface context a number of questions arise concerning the appropriateness of the specification set out in Section 2 of this paper for alphabetic optimality. These concern

(i) Formulation in terms of the $\underset{\sim\sim}{RO}$ region

(ii) Distribution of information over a wider region

(iii) Sensitivity of criteria to size and shape of the $\underset{\sim\sim}{RO}$ region

(iv) Ignoring of bias.

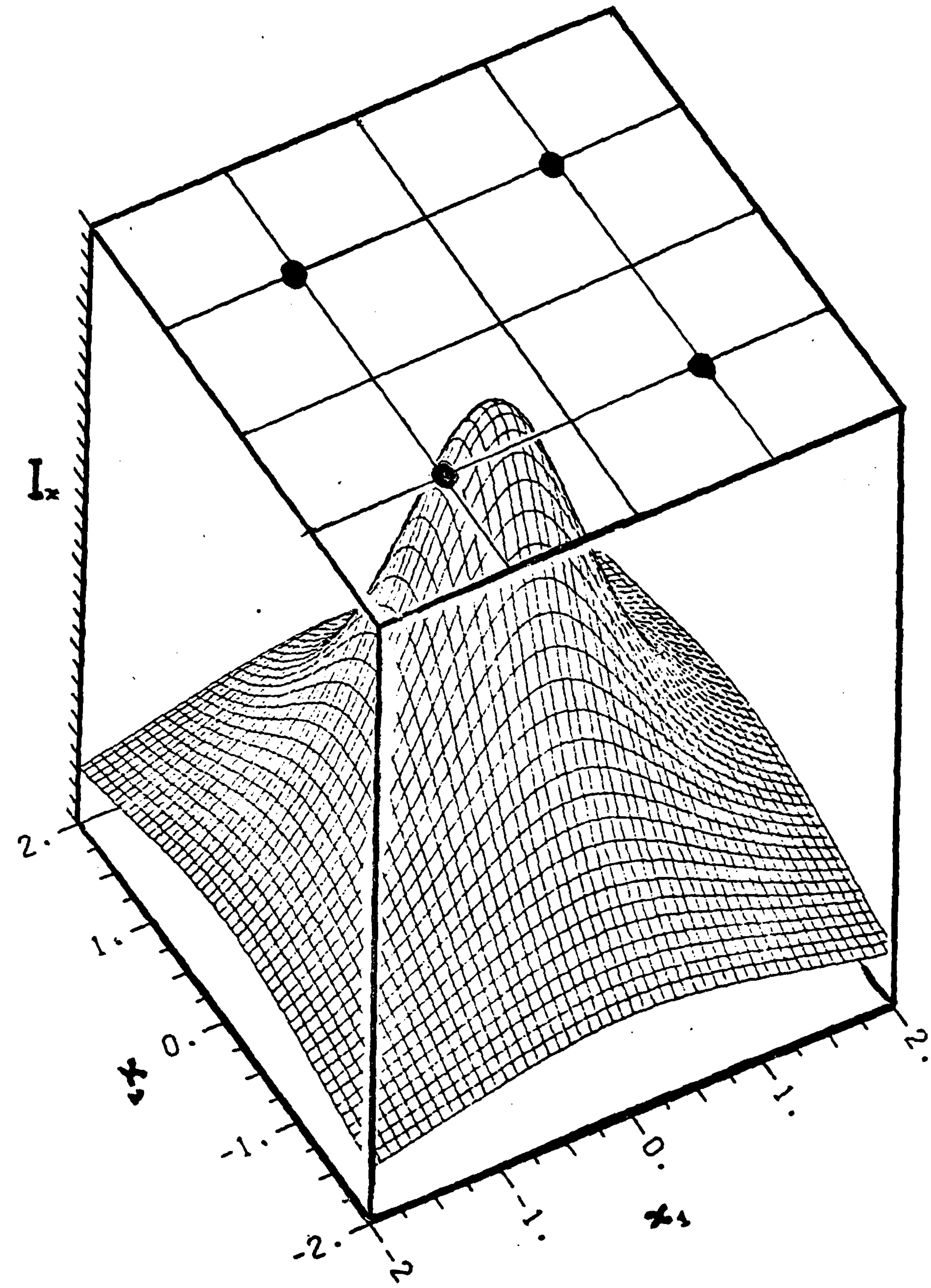


Figure 2(a) Information surface for a 2^2 factorial used as a first order design.

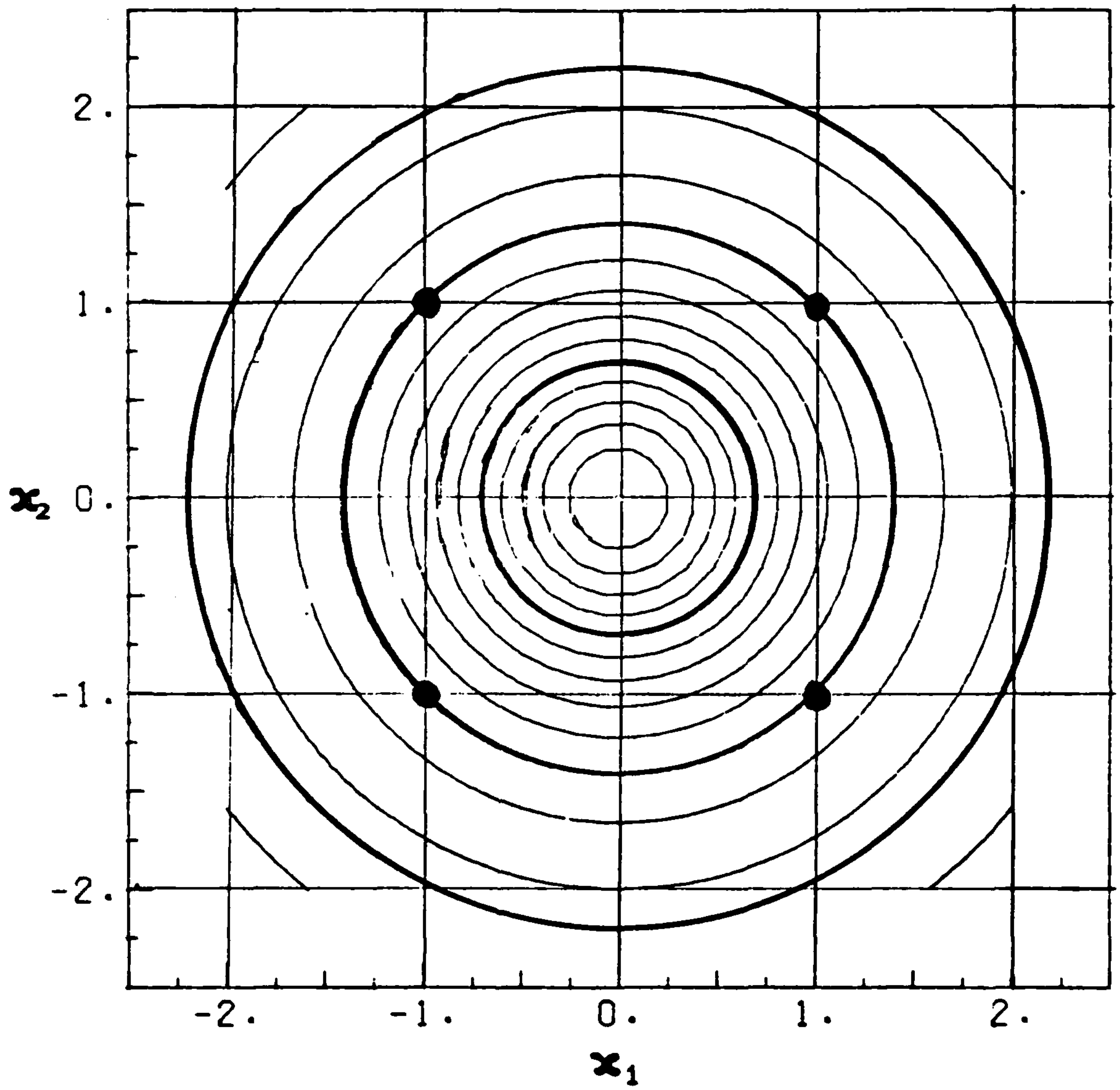


Figure 2(b) Information contours for a 2^2 factorial used as a first order design.

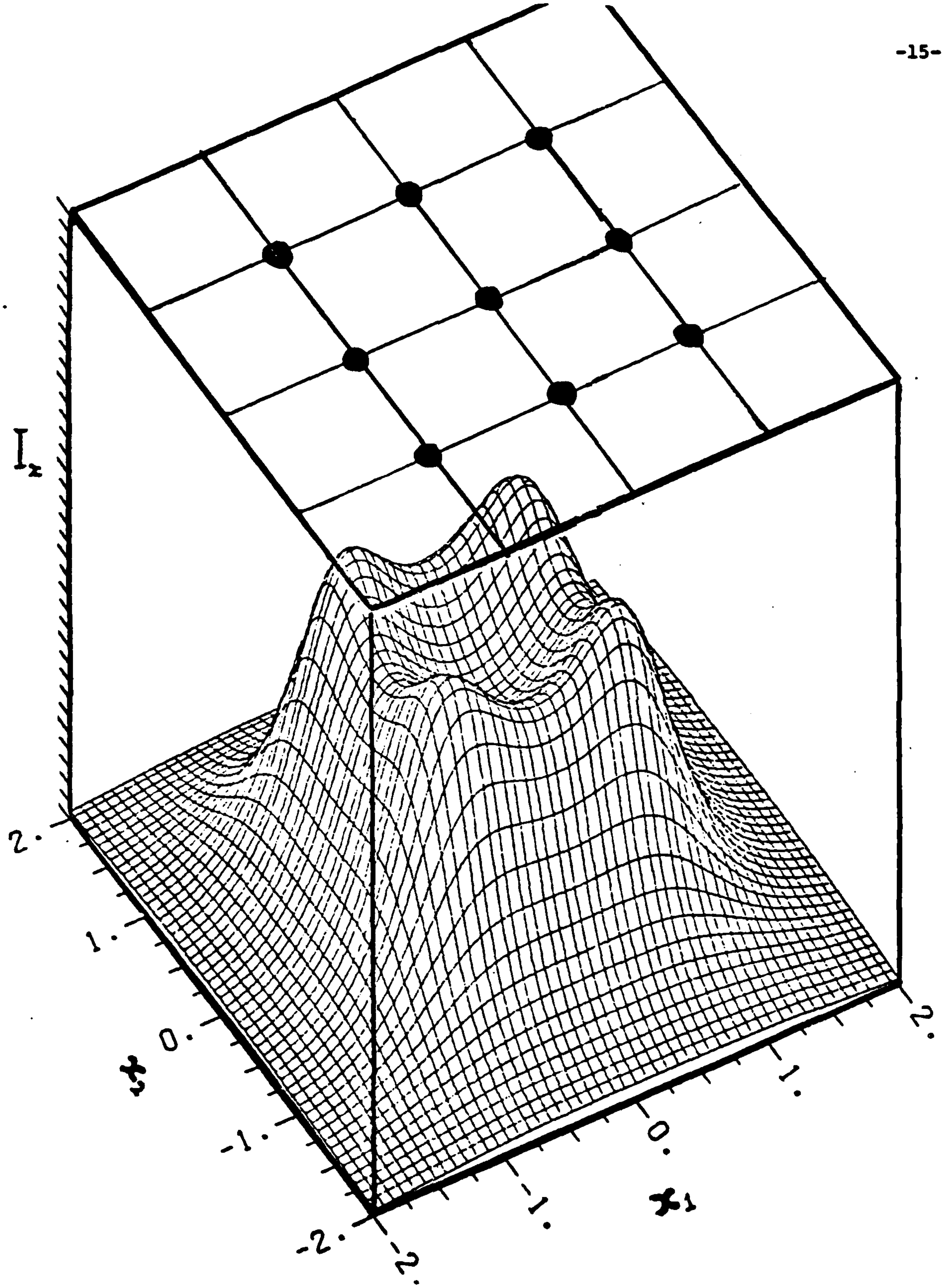


Figure 3(a) Information surface for a 3^2 factorial used as a second order design.

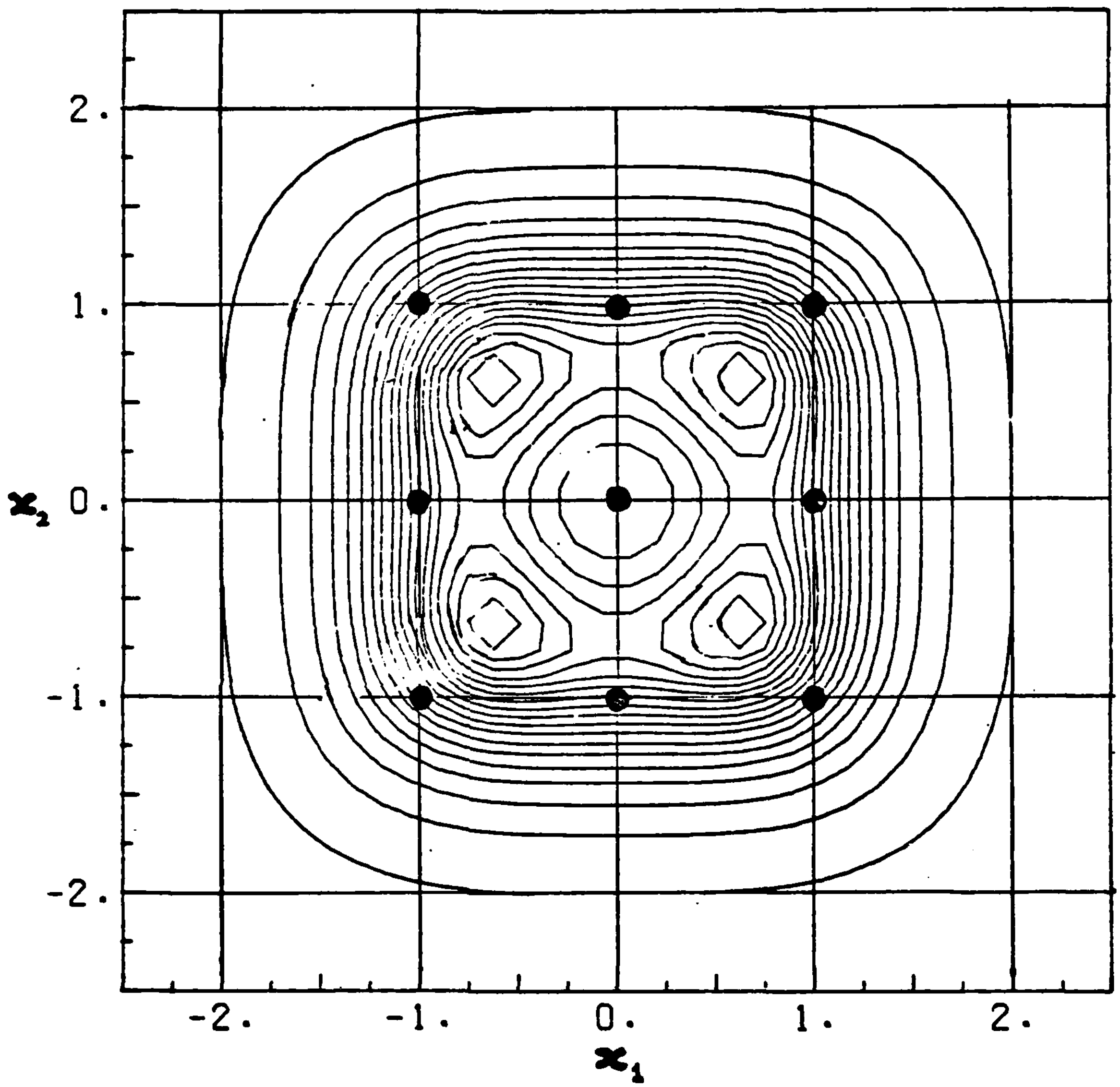


Figure 3(b) Information contours for a 3^2 factorial used as a second order design.

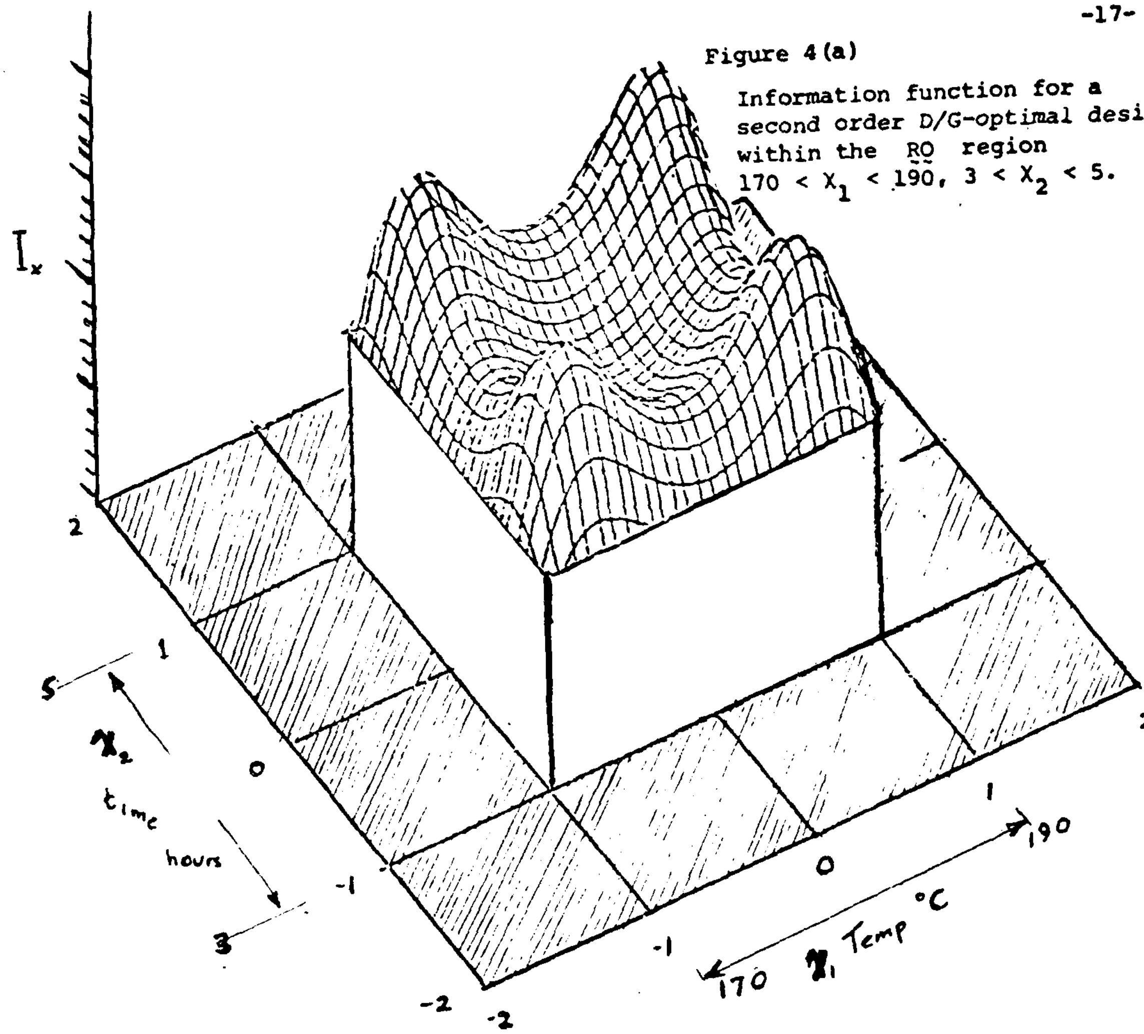


Figure 4(a)

Information function for a second order D/G-optimal design within the $\underset{\sim}{RO}$ region $170 < X_1 < 190$, $3 < X_2 < 5$.

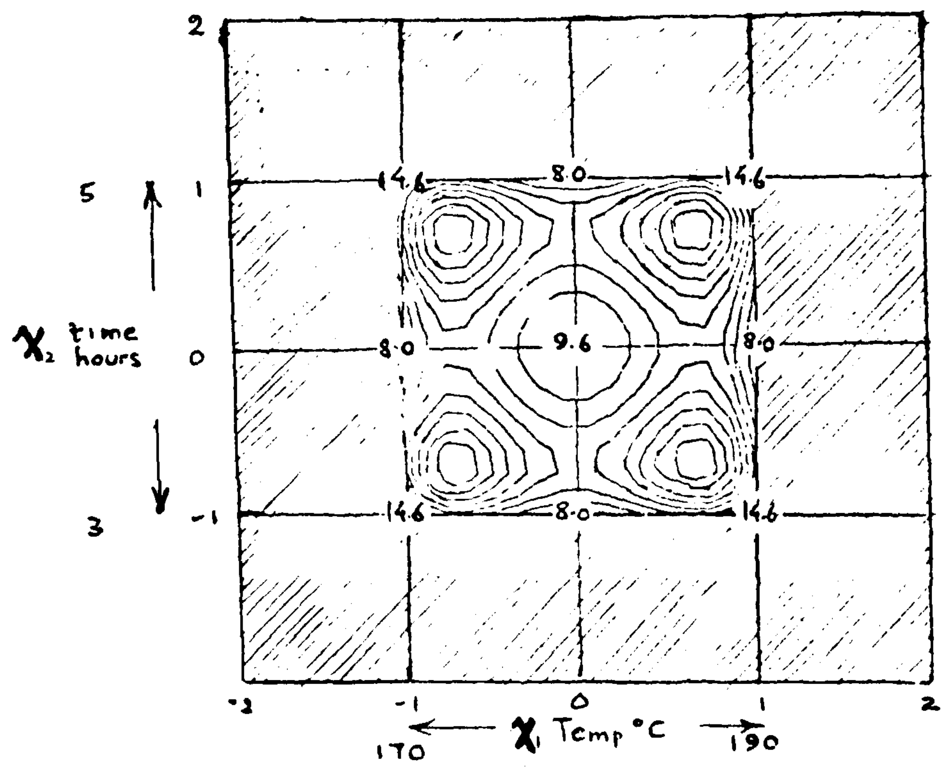


Figure 4(b)

Information contour plot and % distribution of the design measure.

Formulation in terms of the RO region

As has been pointed out, in response surface studies it is typically true that at any given stage of an investigation the current region of interest $\underset{\sim}{R}$ is much smaller than the region of operability $\underset{\sim}{O}$ which is, in any case, usually unknown. In particular, it is obvious that this must be so for any investigation in which we allow the possibility that results of one design may allow progress to a different unexplored region. Consequently I believe that formulation in terms of an $\underset{\sim}{R}\underset{\sim}{O}$ region which assumes that $\underset{\sim}{R}$ and $\underset{\sim}{O}$ are identical is artificial and limiting. In particular, to obtain a good approximation <u>within $\underset{\sim}{R}$</u> one may very well wish to put some experimental points <u>outside $\underset{\sim}{R}$</u> and so long as they are within $\underset{\sim}{O}$ there is *no practical* reason why we should not. Also since typically $\underset{\sim}{R}$ is only vaguely known, we will want to consider the information function over a wider region, as is done for example in Figure 5 for Fedorov's second order D-optimal design. The information function for this design may now be compared over this wider region with that for the 3^2 factorial in Figure 3.

Distribution of information over a wider region

In the response surface context, the coefficients $\underset{\sim}{\beta}$ of a graduating function $\eta_{\underset{\sim}{x}} \doteq \underset{\sim}{x}'\underset{\sim}{\beta}$ acting as they do merely as adjustments to a kind of mathematical french curve are not usually of individual interest except insofar as they affect η, in which case only the G-optimality criterion among those considered is of direct interest. For response surface studies however, it is far from clear how desirable is the property of G-optimality itself.

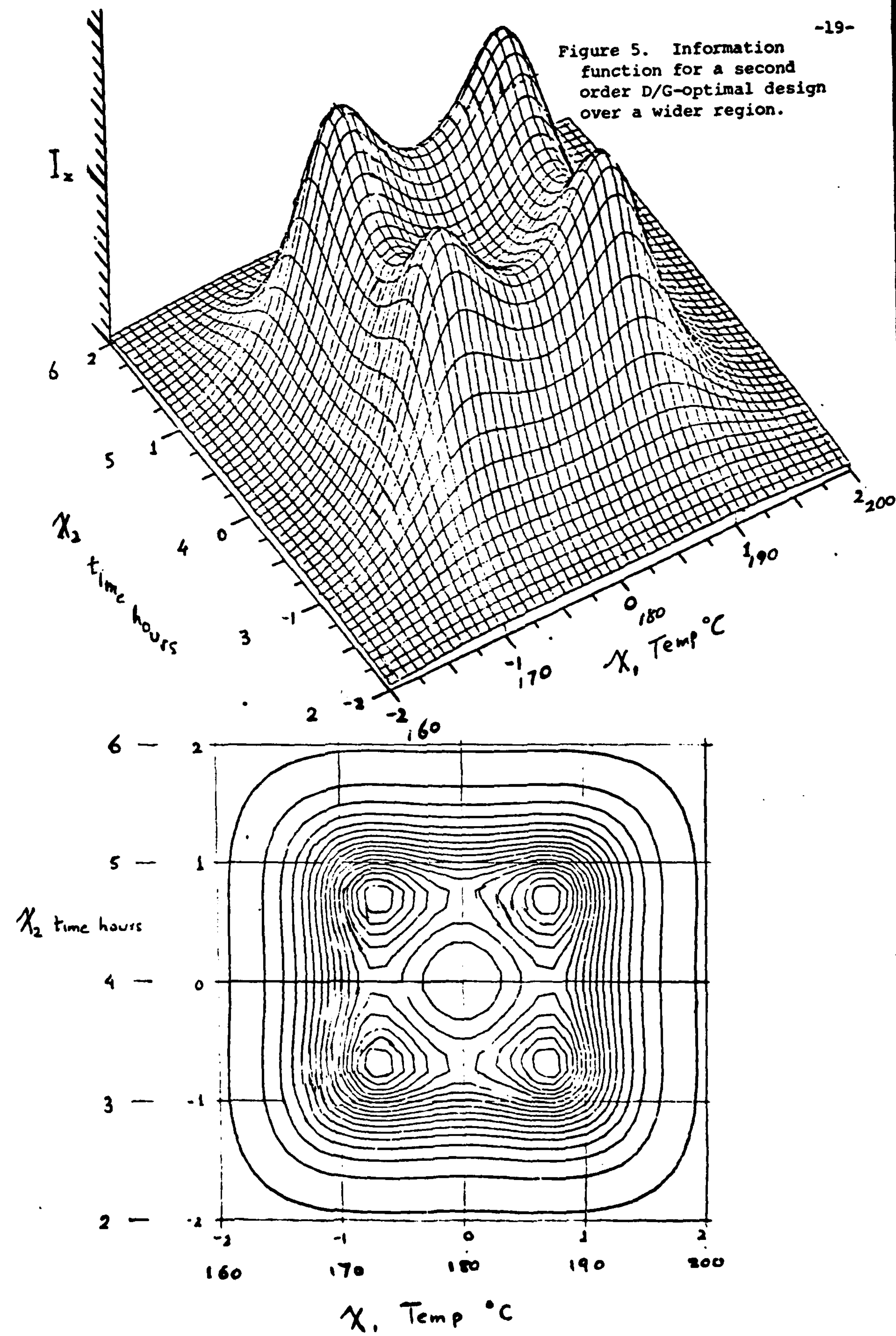
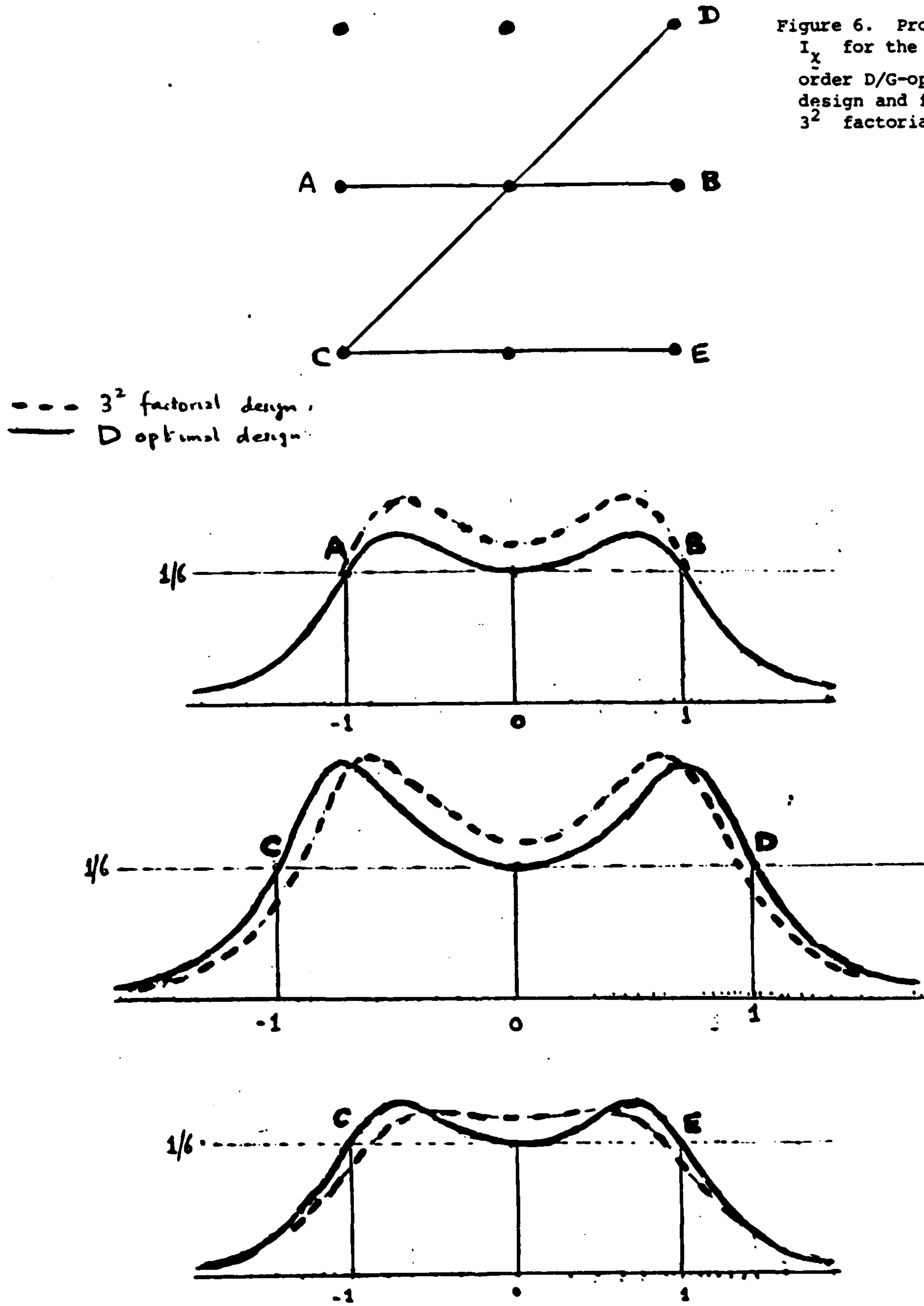


Figure 5. Information function for a second order D/G-optimal design over a wider region.

For instance, the profiles of Figure 6 made by taking sections of the surfaces of Figure 3 and Figure 5 suggest that neither the G/D-optimal design nor the 3^2 design are universally superior one to the other. In some subregions one design is slightly better, and in others the other design is slightly better. Both information functions, and particularly that of the G/D-optimal design, show a tendency to sag in the middle. This happens for the G/D-optimal design because the G-optimality characteristic guarantees that (maximized) minima for $I_{\underset{\sim}{x}}$, each equal to 1/p, occur at every design point, <u>which must include the center point</u>. However, this sagging information pattern of the second order design is not of course a characteristic of the first order design of Figure 2 which is also D/G-optimal but contains no center point. If the idea of the desirability of designs possessing a particular kind of information profile is basic, then it seems unsatisfactory that the nature of that profile should depend so very much on the order of the design. Indeed, the relevance of the minimax criterion which produces G-optimality is arguable. It follows from the Kiefer-Wolfowitz theorem that a second order design for the $(\pm 1, \pm 1)$ region whose information function did not sag in the middle would necessarily not be D-optimal. But as we have seen, D-optimality is only one of many single-valued criteria that might be used in attempts to describe some important characteristic of the $\underset{\sim}{X}'\underset{\sim}{X}$ matrix. Others for example would be A-optimality and E-optimality, and these would yield different information profiles. But I would argue that since the information function <u>itself</u> is the most direct measure of desirability so far as the single issue of variance properties is concerned, our best course is to choose our design directly by picking a suitable information function, and not indirectly by finding some extremum for A, E, D, or other arbitrary criterion.

Figure 6. Profiles of $I_{\tilde{x}}$ for the second order D/G-optimal design and for the 3^2 factorial design.



Sensitivity of criteria to size and shape

In the process of scientific investigation, the investigator and the statistician must do a great deal of guesswork. In matching the region of interest $\underset{\sim}{R}$ and the degree of complexity of the approximating function, they must try to take into account, for example, that a more flexible second degree approximating polynomial can be expected to be adequate over a larger region $\underset{\sim}{R}$ than a first degree approximation. Obviously different experimenters would have different ideas of appropriate locations and ranges for experimental variables. In particular, ranges could easily differ from one experimenter to another by a factor of two or more[5]. In view of this, extreme sensitivity of design criteria to scaling is disturbing[6]. For example, suppose each dimension of a dth order experimental design is increased[7] by a factor c. Then the D criterion is increased by a factor of c^q where

---

[5] Over a sequence of designs, initial bad choices of scale and location would tend to be corrected, of course.

[6] In particular, designs can only be fairly compared if they are first scaled to be of the "same size." But how is size to be measured? It was suggested in [14] that designs should be judged as being of the same size when their marginal second moments $\Sigma(x_{iu} - \bar{x}_u)^2/n$ were identical. This convention is not entirely satisfactory, but will of course give very different results from those which assume design points to be all included in the same region RO. It is important to be aware that the apparent superiority of one design over another will often disappear if the method of scaling the design is changed. In particular this applies to comparisons such as those made by Nalimov et al [40] and Lucas [38].

[7] A measure of efficiency of a design criterion (see for example [3], [17]) is motivated by considering the ratio of the number of runs necessary to achieve the optimal design to the number of runs required for the suboptimal design to obtain the same value of the criterion (supposing fractional numbers of runs to be allowed). In particular for the D criterion, this measure of D-efficiency is $(D/D_{opt})^{1/k}$. Equivalently here, to illustrate scale sensitivity, we concentrate attention on the factor c by which each scale would need to be inflated to achieve the same value of the D criterion.

$$q = \frac{2k(k+d)!}{(k+1)!(d-1)!} .$$

Equivalently a confidence region of the same volume as that for a D-optimal design can be achieved for a design of given D value by increasing the scale for each variable by a factor of $c = (D_{opt}/D)^{1/q}$, thus increasing the volume occupied by the design in the $\underset{\sim}{x}$ space, by a factor $c^k = (D_{opt}/D)^{k/q}$. For example the D value for the 3^2 factorial design of Figure 3 is 0.98×10^{-2} as compared with a D value of 1.14×10^{-2} for the D-optimal design. For (k = 2, d = 2), we find q = 16, and $c = (1.14/0.98)^{1/16} = 1.009$. Thus the same value of D (the same volume of a confidence region for the β's) as is obtained for the D-optimal design would be obtained from a 3^2 design if each side of the square region were increased by less than 1%. Equivalently, the area of the region would be increased by less than 2%. Using the scaling that was used in Figure 4 for illustration, we should have to change the temperature by 20.18 °C instead of 20 °C, and the time by two hours and one minute instead of two hours, for the 3^2 factorial to give the same D value as the D/G-optimal design. Obviously no experimenter can guess to anything approaching this accuracy what are suitable ranges over which to vary these factors.

Obviously choice of region and choice of information function are closely interlinked. For example, <u>any</u> set of N = k+1 points in k-space which have no coplanarities is obviously a D-optimal first order design for <u>some</u>[8] ellipsoidal region. Furthermore the information function for a design of order d is a smooth function whose harmonic average over the n experimental points (which can presumably be regarded as representative of the region of

---

[8] Namely for that region enclosed within the information contour $I_{\underset{\sim}{x}} = 1/p$ which must pass through all the k+1 experimental points.

interest) is always 1/p wherever we place the points. Thus the problem of design is not so much a question of choosing the design to increase total information as spreading the total information around in the manner desired.

Rotatable Designs

A route for simplication different from alphabetic optimality occurs when, after suitable transformation of the inputs $\underset{\sim}{\chi}$ to standardized variables $\underset{\sim}{x}$ nothing is known about the orientation in the $\underset{\sim}{x}$ space of the response surface we wish to study. It was argued by Box and Hunter [11] that we should then employ designs having the property that the variance of $\hat{y}$ is a function only of $\rho = (\underset{\sim}{x}'\underset{\sim}{x})^{1/2}$ so that

$$V_{\underset{\sim}{x}} = V_{\rho} \quad \text{and} \quad I_{\underset{\sim}{x}} = I_{\rho} \quad .$$

For a first order design, rotatability implies orthogonality and vice versa, and completely decides the information function. For second and higher order designs, a requirement of rotatability fixes many moment properties of the design, but V_ρ and hence I_ρ are still to some extent at our choice, and can be changed by changing certain moment ratios [11]. In particular, for a second order design, V_ρ depends on the single moment ratio $\lambda = (n/3)\Sigma x_i^4/(\Sigma x_i^2)^2$. For illustration, Figure 7 shows the information function for a second order rotatable design with $\lambda = .75$ consisting of 8 points arranged in a regular octagon with 4 points at the center.

The truth seems to be that at any particular phase of an investigation the scientific decision that most contributes to the outcome of that phase is the choice of the current region of interest (involving choice of variables, locations, ranges, and transformations) -- this is a choice that does not really involve statistics. After this decision is made, (and given the assumption that the model fits perfectly so that only the variance properties

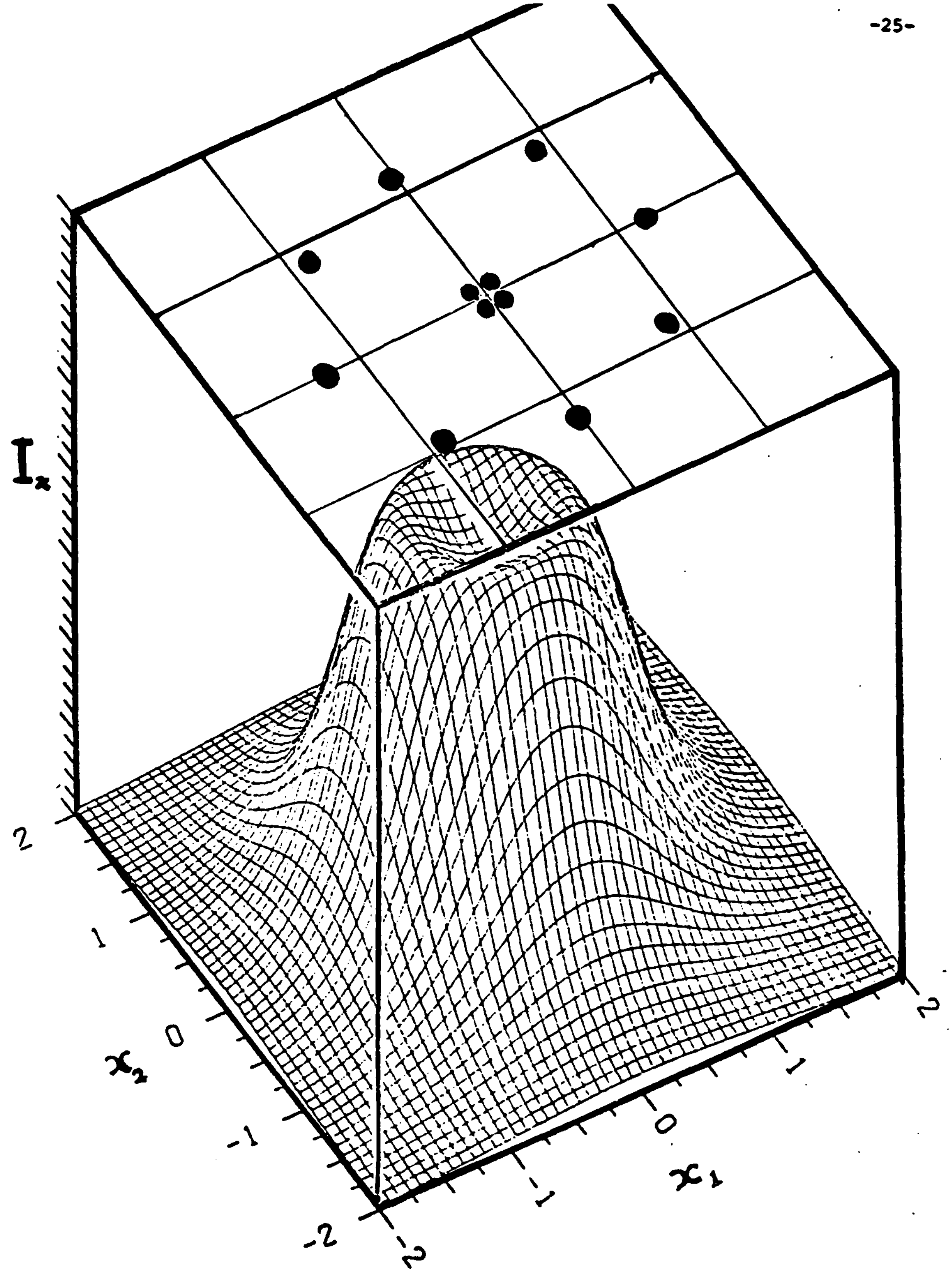


Figure 7(a) Information function for a second order rotatable design consisting of 8 points on a circle plus 4 center points.

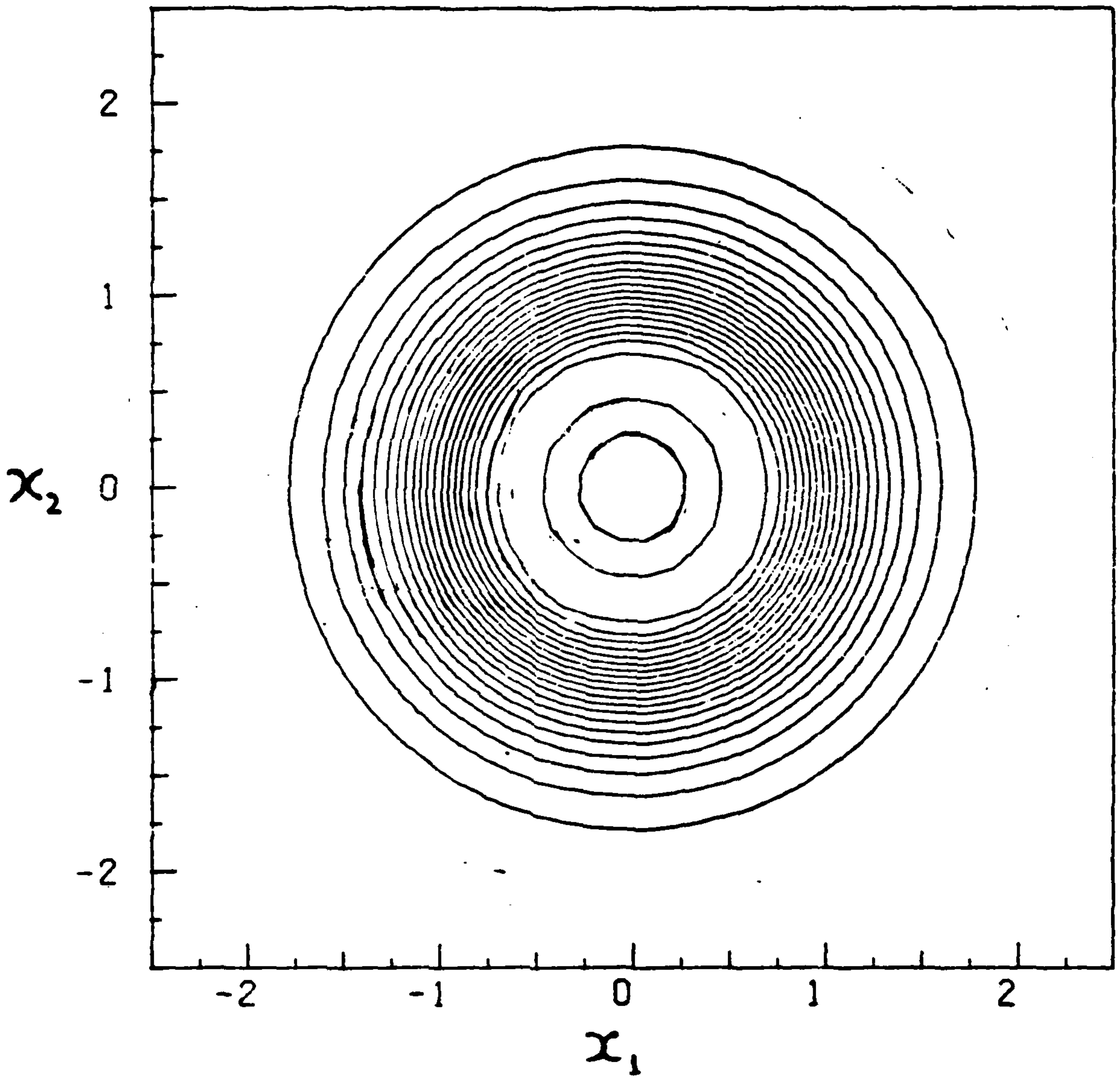


Figure 7(b) Information contours for the same design.

of the design are of interest) any set of experiments that cover this region in some reasonably uniform way is likely to do quite well. I cannot see that the various optimality criteria are particularly relevant to this choice, although there would certainly be no harm in considering them, together with many other factors briefly discussed later.

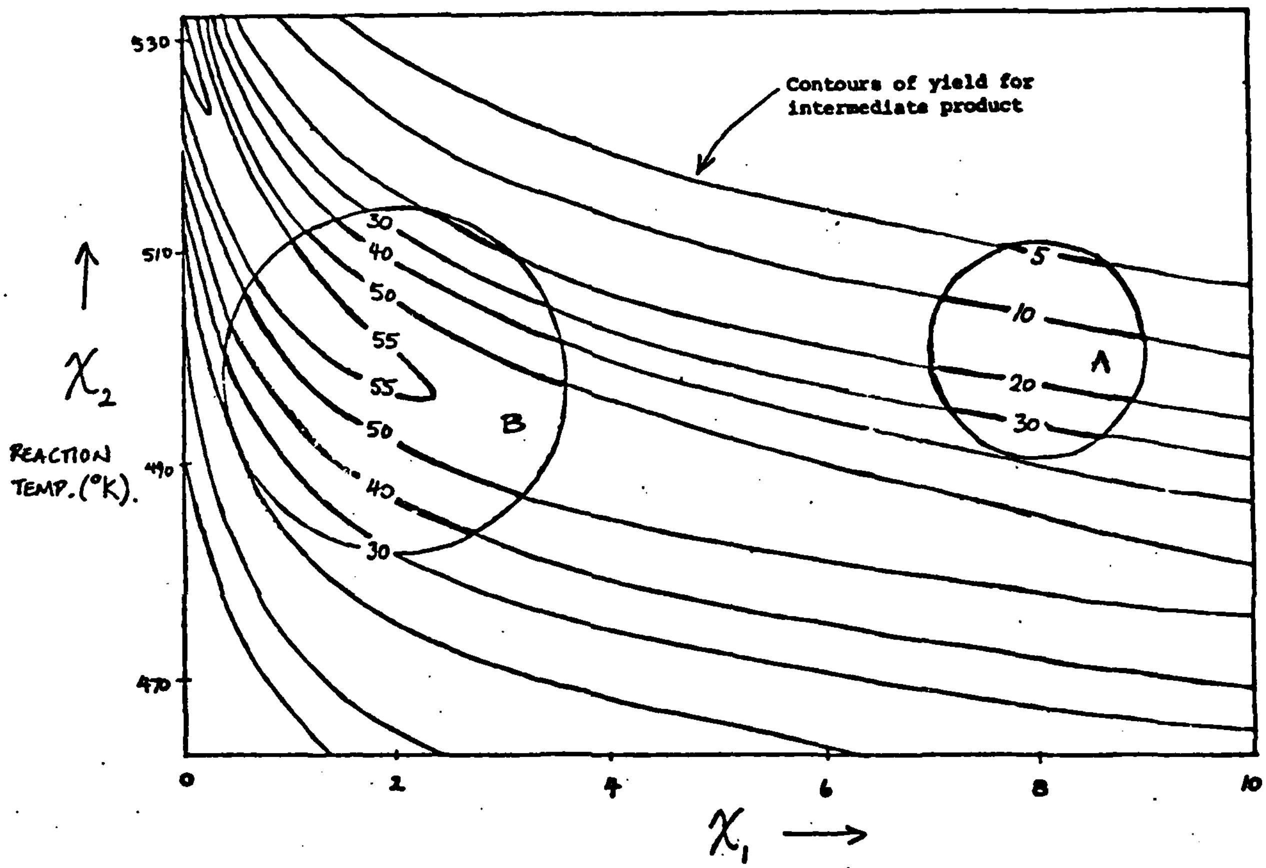
Ignoring of bias

All models are wrong; some models are useful. This aphorism is particularly true for empirical functions such as polynomials that make no claim to do more than locally graduate the true function. For chemical examples some idea of the adequacy of such approximations can be gained by studying surfaces produced by chemical kinetic models. An example[8] taken from [10] is shown in Figure 8. See also [15].

One conclusion I reached from many such studies was that approximations would not need to be very good for response surface methods to work. Thus within region A of Figure 8 the locally monotonic function could be crudely approximated by a plane which could indicate a useful path of ascent. Also valuable information might be obtained about a ridge such as that in region B, even though the underlying surface was not exactly quadratic. Notice however

---

[8] This surface was generated (see [10] for details) by considering the yield of the product B in a consecutive reaction $A \xrightarrow{k_1} B \xrightarrow{k_2} C$ following first order kinetics with temperature sensitivity given by the Arrhenius relation $\ln k_i = \ln \alpha_i + \beta_i/T$, where temperature T is measured in degrees Kelvin, using plausible values for the constants $\alpha_1, \alpha_2, \beta_1, \beta_2$.



REACTION TIME (HOURS)

Figure 8. Contours of a theoretical response surface in reaction time and reaction temperature for a first order consecutive reaction, with plausible values substituted for kinetic constants.

that in the light of such examples any theory of experimental design which depended on the exactness of such approximations should be regarded with some skepticism.

5. TAKING ACCOUNT OF BIAS

If $\hat{y} = \underset{\sim}{x}'\hat{\underset{\sim}{\beta}}$ is the fitted value using the empirical approximation, then its total error ε is

$$\hat{y} - \eta = \{\hat{y} - E(\hat{y})\} + \{E(\hat{y}) - \eta\}$$

$$\varepsilon \quad = \quad \varepsilon_V \quad + \quad \varepsilon_B \ .$$

Thus the error ε contains a random part ε_V and a systematic, or bias, part ε_B, and we must expect that ε_B will <u>not</u> be negligible. Since all the theory previously discussed makes the assumption that ε_B is zero, we must consider whether the resulting designs are robust to this kind of discrepancy. The optimality criteria discussed earlier which assume the response function to be exact usually produce a substantial proportion of experimental points on the boundary of $\underset{\sim}{R}\underset{\sim}{O}$. In the context of possible bias, this is not reassuring, since it is at these points that the approximating function will be most strained.

The explicit recognition that bias will certainly be present does however seem to provide a more rational means for approaching the scaling problem ([6], [7]). To see this, consider again the formulation given earlier in terms of a region of interest $\underset{\sim}{R}$ and a larger region $\underset{\sim}{O}$ of operability. If we were to assume (unrealistically) that the approximation remained exact

however widely the points were spread, and if some measure of variance reduction were the only consideration, then to obtain most accurate estimation within $\underset{\sim}{R}$, the size of the design would have to be increased to the boundaries of the operability region $\underset{\sim}{O}$. But in fact of course the wider the points were spread, the less applicable would be the approximating function, and the bigger the bias error. This suggests that we should seek restriction of the spread of the experimental points not by artificial limitation to some region $\underset{\sim}{RO}$, but by balancing off the competing requirements of variance on the one hand, which is reduced as the spread of the points is increased, and bias on the other hand, which is increased as the spread of the points is increased.

The mean square error associated with estimating $\eta_{\underset{\sim}{x}}$ by $\hat{y}_{\underset{\sim}{x}}$ standardized for the number, n, of design points and the error variance σ^2, can be written as the sum of a variance component and a squared bias component

$$n \cdot E(\hat{y}_{\underset{\sim}{x}} - \eta_{\underset{\sim}{x}})^2/\sigma^2 = n \cdot V(\hat{y}_{\underset{\sim}{x}})/\sigma^2 + n\{E(\hat{y}_{\underset{\sim}{x}}) - \eta_{\underset{\sim}{x}}\}^2/\sigma^2 ,$$

or

$$M_{\underset{\sim}{x}} = V_{\underset{\sim}{x}} + B_{\underset{\sim}{x}} .$$

For illustration, an example is taken from a forthcoming book with N.R. Draper and J.S. Hunter [10]. Figure 9 shows a situation as it might exist for a single variable when a straight line approximating function is to be used. The diagram shows what might be the true underlying function which would of course be obscured by experimental error. Suppose the region of interest $\underset{\sim}{R}$ is scaled so that $-x_0 < x < x_0$ and in particular consider the two designs

(a) (-2/3, 0, 2/3) and (b) (-4/3, 0, 4/3).

One way [6] to obtain overall measures of variance and squared bias over any specified region of interest $\underset{\sim}{R}$ is by averaging $V_{\underset{\sim}{x}}$ and $B_{\underset{\sim}{x}}$ over $\underset{\sim}{R}$ to provide the quantities

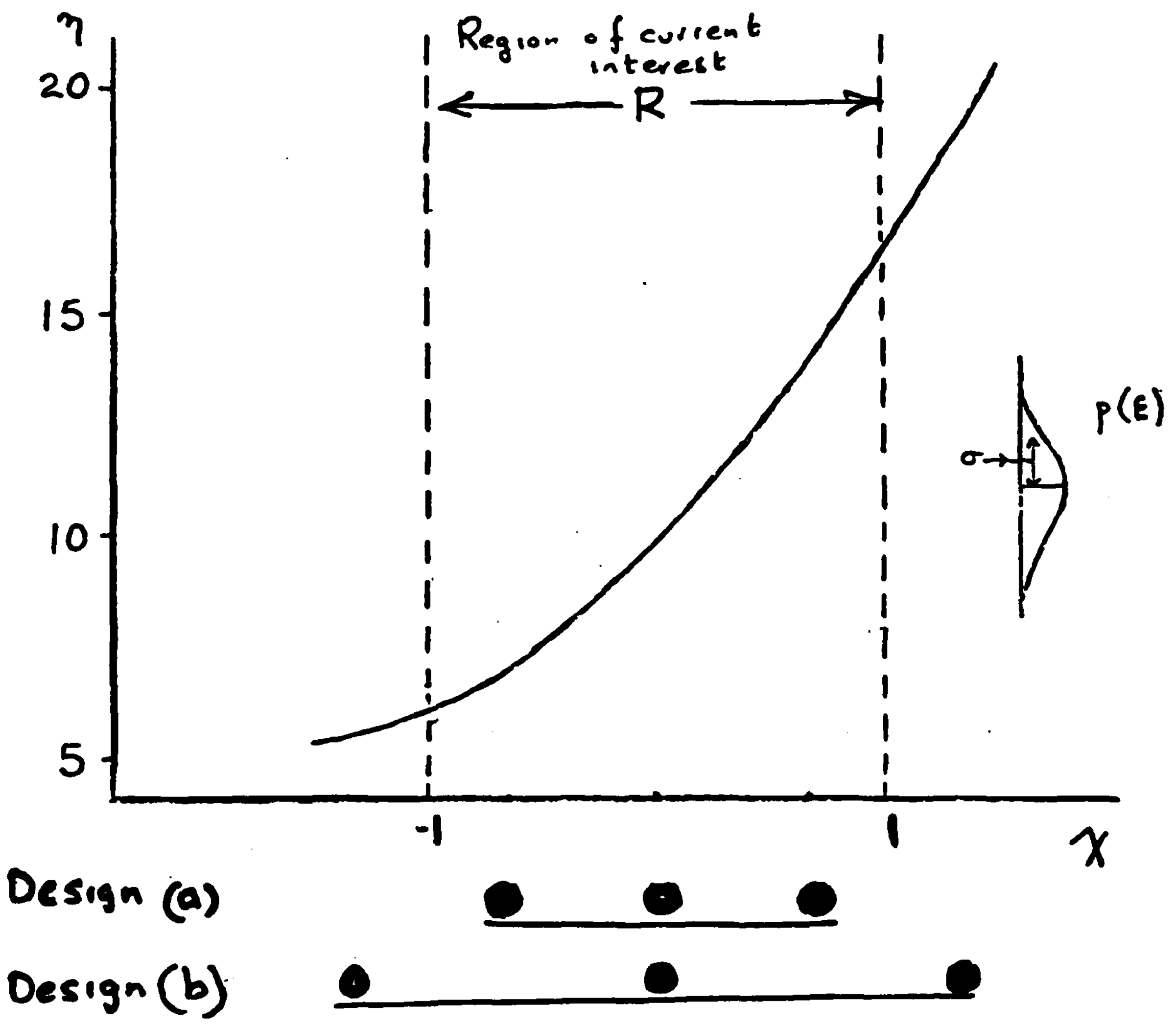


Figure 9. Two possible designs for fitting a straight line over a region of interest $\underset{\sim}{R}$ $-1 < \chi < 1$.

$$V = \int_{\underset{\sim}{R}} V_{\underset{\sim}{x}} d\underset{\sim}{x} \Big/ \int_{\underset{\sim}{R}} d\underset{\sim}{x} \quad \text{and} \quad B = \int_{\underset{\sim}{R}} B_{\underset{\sim}{x}} d\underset{\sim}{x} \Big/ \int_{\underset{\sim}{R}} d\underset{\sim}{x} .$$

Denoting the integrated (over $\underset{\sim}{R}$) mean square error by M, we can then write

$$M = V + B.$$

For the previous example, V, B, and M are plotted against x_0 in Figure 10. We see how V becomes very large if the spread of the design is made very small, while if the design is made very large, V slowly approaches its minimum value of unity. The average squared bias B, on the other hand, has a minimum value when x_0 is about 0.7, and increases for larger or smaller designs. A rather flat minimum for M = V + B occurs near $x_0 = 0.79$. Thus in this manner the design which minimizes average mean squared error M is not very different from the design which minimizes average squared bias B, but is extremely different from that which minimizes average variance V.

Choice of alternative model

A difficulty in all this is that in practice we do not know the nature of the true function $\eta_{\underset{\sim}{x}}$. Progress may be made however by supposing that $\eta_{\underset{\sim}{x}}$ is to some satisfactory approximation represented by a polynomial model of higher degree d_2. Suppose then that a polynomial model of degree d_1 is fitted to n data values to give

$$\hat{\underset{\sim}{y}}_{\underset{\sim}{x}} = \underset{\sim}{X}_1 \underset{\sim}{b}_1$$

while the true model is in fact a polynomial of degree d_2, so that

$$\underset{\sim}{\eta}_{\underset{\sim}{x}} = \underset{\sim}{X}_1 \underset{\sim}{\beta}_1 + \underset{\sim}{X}_2 \underset{\sim}{\beta}_2 .$$

We also need to know something about the relative magnitudes of systematic and random errors that we could expect to meet in practical cases. It was argued in [6] that an investigator might typically employ a

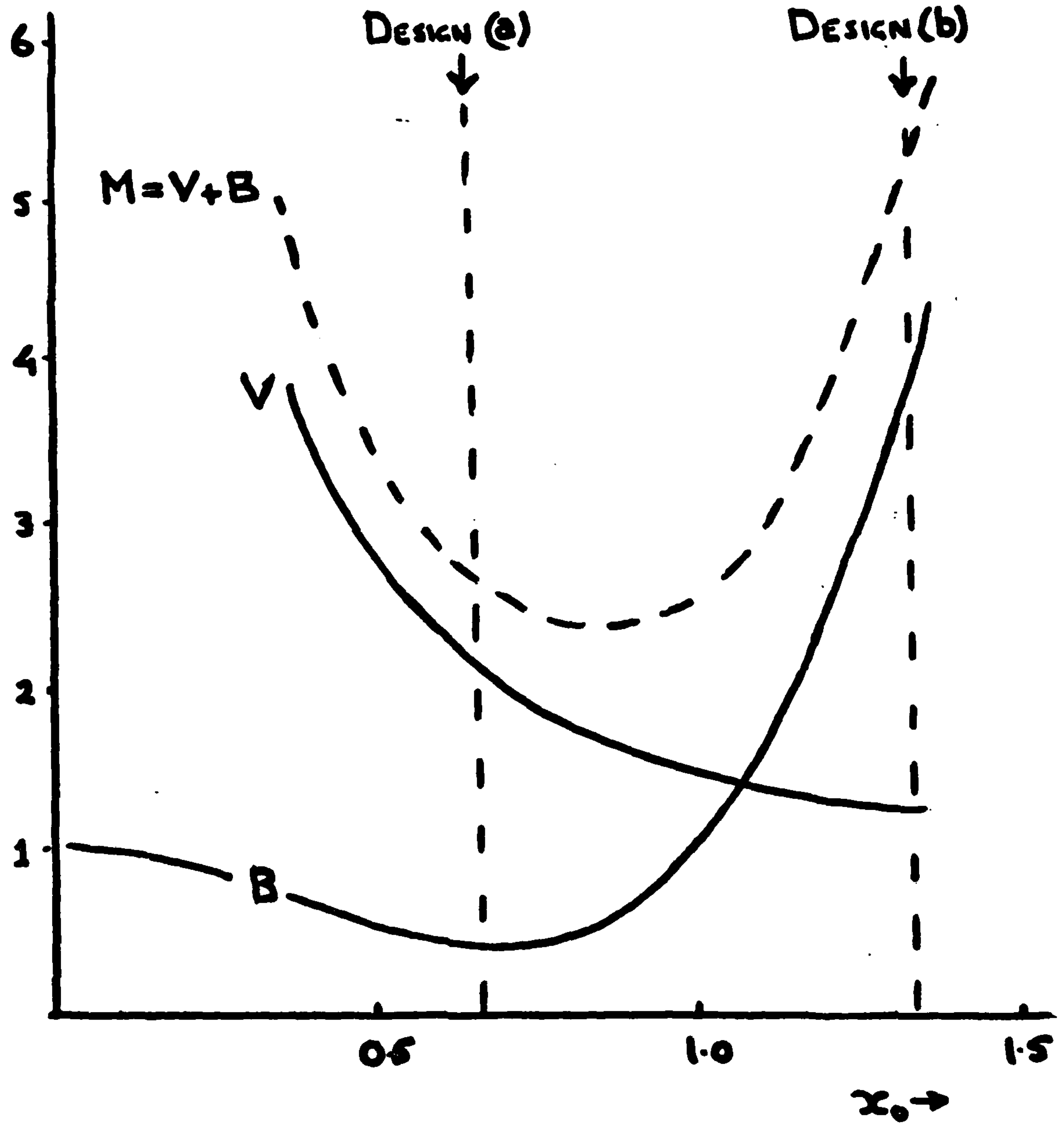


Figure 10. The behavior of integrated variance V, integrated squared bias B, and their total M = V + B, for three-point designs $(-x_0,0,x_0 0)$.

fitted approximating function such as a straight line when he believed that the average departure from the truth induced by this approximating function was no worse than that induced by the process of fitting. This would suggest that the experimenter would tend to choose the size of his region $\underset{\sim}{R}$, and the degree of his approximating function in such a way that the integrated random error and the integrated systematic error were about equal. Thus we might suppose that a situation of particular interest is that where B is roughly equal to V. Examples that we studied seemed to show that designs that minimized M with the constraint $V \approx B$ were close to those which minimized B. Consequently we suggested that, if a simplification were to be made in the design problem, it might almost be better to ignore the effects of sampling variation rather than those of bias.

However this may be, there seems no doubt that, in making a table of useful designs, a component in our thinking should be the characteristics of the designs which minimized squared bias against feared alternatives. As a factor in our final choice, this should certainly receive as much attention as the indications supplied by, say, D-optimality.

For illustration particular examples of designs in three dimensions which minimize integrated squared bias when $\underset{\sim}{R}$ is a sphere of unit radius are shown in Figure 11(a) for $d_1 = 1$ and $d_2 = 2$ (a first order design robust to second order effects) and in Figure 11(b) for $d_1 = 2$ and $d_2 = 3$ (a second order design robust to third order effects). The former is the familiar 2^3 factorial scaled so that the points are 0.71 units from the center. The latter is a rotatable composite design with "cube" points at a distance 0.86 from the center, and "star" points at a distance 0.83 from the center.

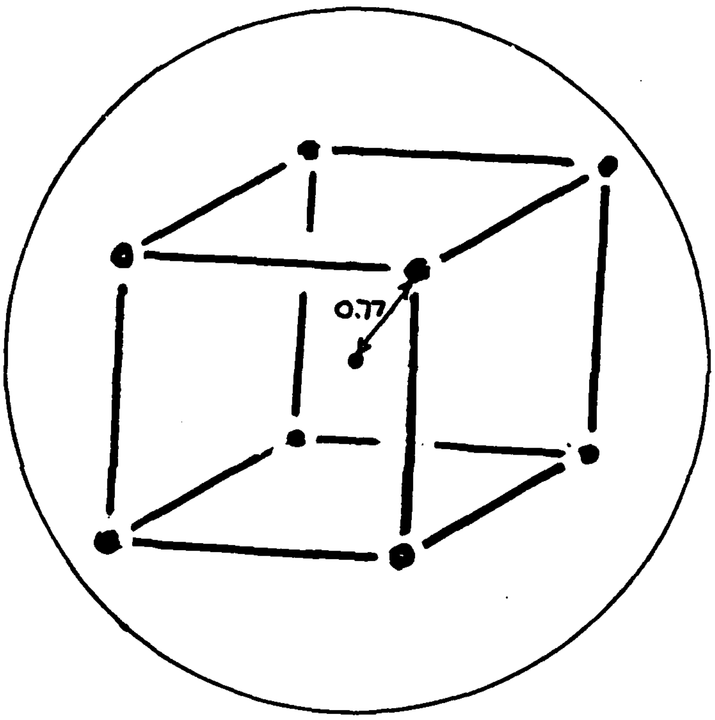


Figure 11(a) A first order (two-level factorial) design in three factors which minimizes squared bias from second order terms when the region of interest is a sphere of unit radius.

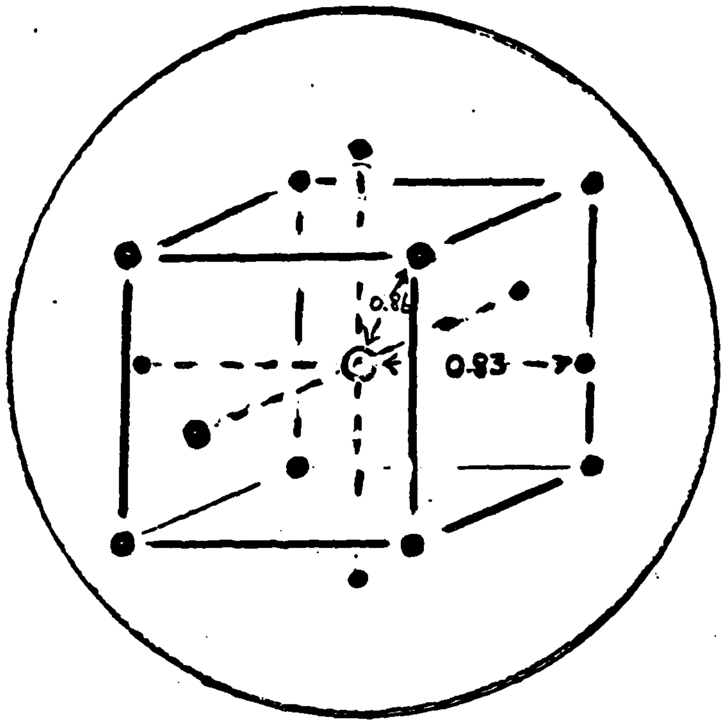


Figure 11(b) A second order composite rotatable design which minimizes squared bias from third order terms when the weight function is uniform over a spherical region of interest of unit radius.

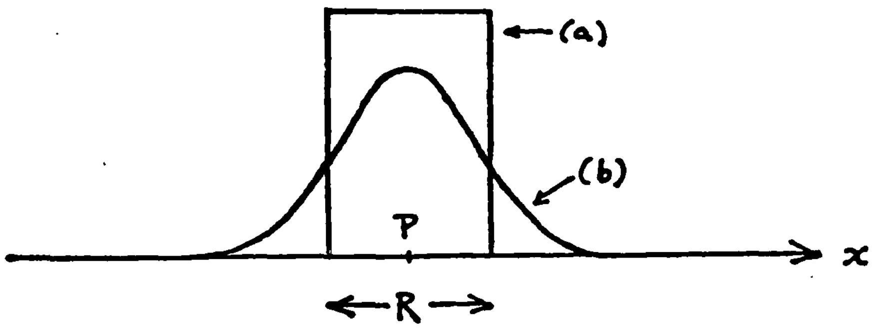


Figure 12: Two possible weight functions for $k = 1$:
(a) "Uniform over R" type indicating uniform interest over R, no interest outside R;
(b) Normal Distribution shape, giving greater weight to points nearer P.

Obviously in practice because of the inevitable inexactness of choosing scales exact dimensions of the designs should not be taken too seriously, but these examples illustrate the fact that as soon as we take account of bias, design points are not chosen on the boundary of R.

Choice of designs which minimize bias

Before considering the problem of choosing minimum bias designs it is desirable to generalize slightly the previous formulation. Although it avoids limiting the location of the design points in an artificial way the idea of a region of interest $\underset{\sim}{R}$ within a larger operability region $\underset{\sim}{O}$ is still not entirely satisfactory because it implies that we have equal interest at all points within $\underset{\sim}{R}$. A more general formulation [7] which subsumes that we have been discussing employs a weight function $w(\underset{\sim}{x})$ which extends over the operability region $\underset{\sim}{O}$ so that $\int_{\underset{\sim}{O}} w(\underset{\sim}{x})d\underset{\sim}{x} = 1$. The weighted mean square error M can now be split into a weighted variance part V and a weighted squared bias part B so that again $M = V + B$, with

$$M = \int_{\underset{\sim}{O}} w(\underset{\sim}{x}) E\{\hat{y}(\underset{\sim}{x}) - \eta(\underset{\sim}{x})\}^2 d\underset{\sim}{x}$$

$$V = \int_{\underset{\sim}{O}} w(\underset{\sim}{x}) E\{\hat{y}(\underset{\sim}{x}) - E(\hat{y}(\underset{\sim}{x}))\}^2 d\underset{\sim}{x}$$

$$B = \int_{\underset{\sim}{O}} w(\underset{\sim}{x}) \{E(\hat{y}(\underset{\sim}{x}) - \eta(\underset{\sim}{x})\}^2 d\underset{\sim}{x} .$$

Two possible weight functions for $k = 1$ [20] are shown in Figure 12.

Suppose as before the fitted function is a polynomial $\underset{\sim}{X}_1\underset{\sim}{b}_1$ of degree d_1 while the true model is a polynomial $\underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2$ of degree d_2 and define moment matrices for the design and for the weight function by

$$\underset{\sim}{M}_{11} = n^{-1}\underset{\sim}{X}_1'\underset{\sim}{X}_1, \qquad \underset{\sim}{M}_{12} = n^{-1}\underset{\sim}{X}_1'\underset{\sim}{X}_2$$

$$\underset{\sim}{\mu}_{11} = \int_{\underset{\sim}{O}} w(\underset{\sim}{x})\underset{\sim}{x}_1\underset{\sim}{x}_1'd\underset{\sim}{x}, \qquad \underset{\sim}{\mu}_{12} = \int_{\underset{\sim}{O}} w(\underset{\sim}{x})\underset{\sim}{x}_1\underset{\sim}{x}_2'd\underset{\sim}{x}.$$

Then [6] a necessary and sufficient condition for the squared bias B to be minimized is that

$$\underset{\sim}{M}_{11}^{-1}\underset{\sim}{M}_{12} = \underset{\sim}{\mu}_{11}^{-1}\underset{\sim}{\mu}_{12}$$

and hence a sufficient condition is that all the moments of the design up to and including order $d_1 + d_2$, are equal to all the corresponding moments of the weight function.

6. SOME OTHER CONSIDERATIONS IN DESIGN CHOICE

There is insufficient space to discuss here all of the items in Table 1 that, in one circumstance or another, it might be necessary to take into account, but mention will be made of a few.

Lack of Fit (iii), Sequential Assembly (vi), Blocking (v), Estimation of Error (vii), Transformation Estimation (iv)

While the adequacy of a particular approximating function to explore a region of current interest is always to some extent a matter of guesswork, simple approximations requiring fewer runs for their elucidation will usually be preferred to more complicated ones. This leads to a strategy of building up from simpler models, rather than down from more complicated ones. A

practical procedure is then: to employ the simplest approximating function which it is hoped may be adequate; to allow for checking its adequacy of fit (see also [1], [2], [6], and [19]); to switch to a more elaborate approximating function when this appears necessary. The implication for designs is (a) that they should provide for checking model adequacy, (b) that they should be capable of sequential assembly -- a design of order d should be augmentable to one of order d + 1, (c) since conditions may change slightly from one set of runs to another, especially affecting level, the pieces of the design should form orthogonal blocks.

For illustration, Figure 13 shows the sequential assembly of a design arranged in three orthogonal blocks, each of six runs, labeled I, II, and III. Block I is a first order design but also provides a check for overall curvature (obtained by contrasting the average response of the center points with the average response on the cube). A single contrast of the center response is available as a gross check on previous information about experimental error. If after analyzing the results from Block I there are doubts about the adequacy of a first degree polynomial model, Block II may be performed. It uses the complementary simplex, and the two parts together form a first order design (I+II) with much greater ability to detect lack of fit due to second order terms provided by additional orthogonal contrasts estimating the two-factor interactions. The addition of Block III produces a composite design (I+II+III) which allows a full second degree approximating equation to be fitted if this appears to be desirable. The complete design also provides orthogonal checking contrasts for lack of quadraticity in each of the three directions ([9], [12]). These contrasts can also be regarded as checking the need for transformation in each of the X's. Finally if it were

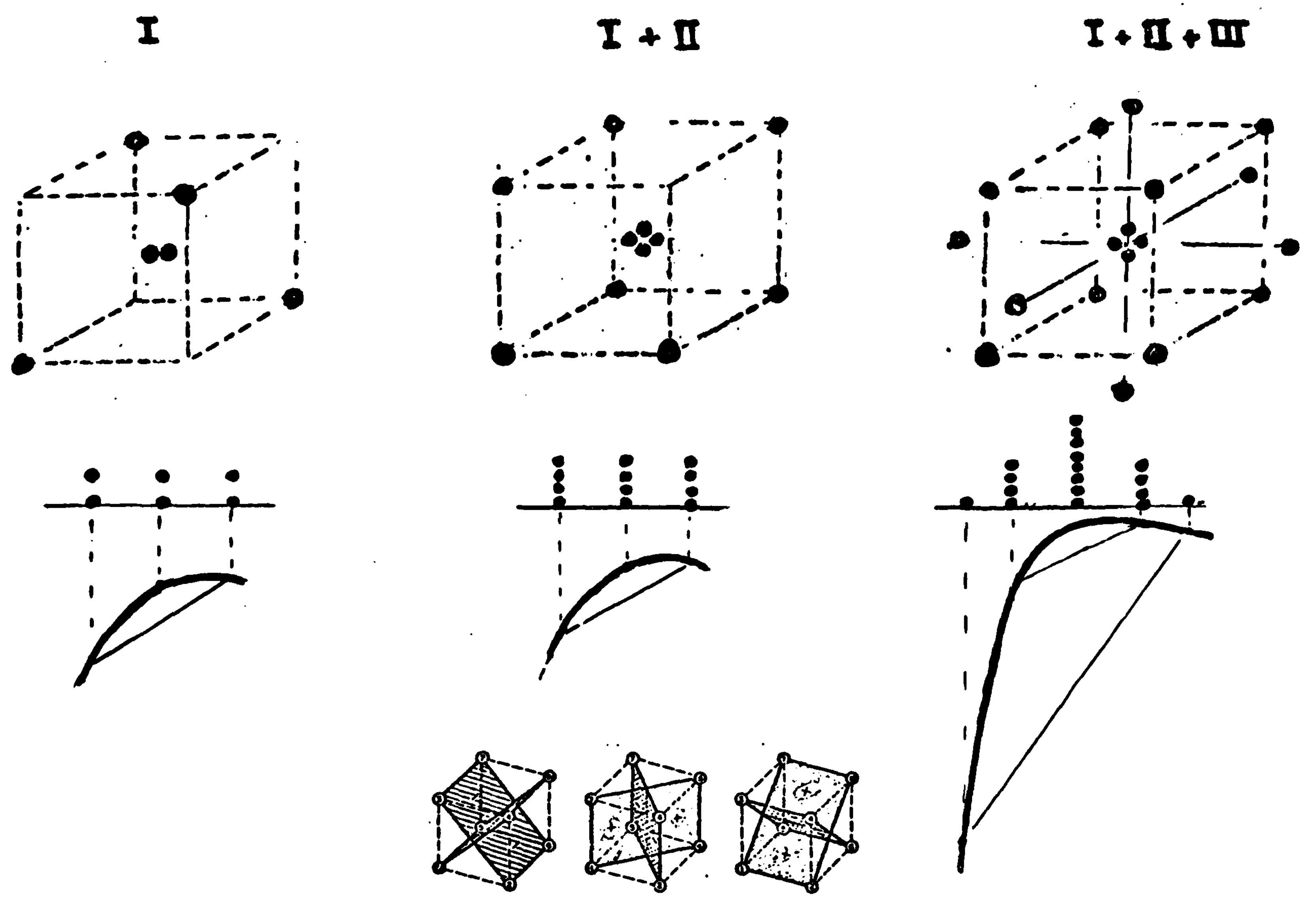


Figure 13. An example of sequential assembly, showing checks of linearity and quadraticity.

decided that more information about experimental error was desirable, the replication of the star in a further Block IV could furnish this, and also provide some increase in the robustness of the design to wild observations.

Robustness

Approaches to the robust design of experiments have been recently reviewed by Herzberg [28]; see also [29]. In particular, Box and Draper [8] suggested that the effects of wild observations could be minimized by making $r = \Sigma r_{uu}^2$ small, where $\underset{\sim}{R} = \{r_{tu}\} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$. This is equivalent to minimizing $\Sigma r_{uu}^2 - p^2/n = \mathrm{Var}\{V(\hat{y})\}$ which takes the value zero when $V(\hat{y}_u) = p/n$ $(u = 1,2,\ldots,n)$. Thus G-optimal designs are optimally robust in this sense.

Size of the experimental design

A good experimental design is one which focuses experimental effort on what is judged important in the particular current experimental context. Suppose that, in addition to estimating the p parameters of the assumed model form, it is concluded that $f > 0$ contrasts are needed to check adequacy of fit, $b > 0$ further contrasts for blocking, and that an estimate of experimental error is needed having $e > 0$ degrees of freedom. To obtain independent estimates of all items of interest we then require a design containing at least $p + f + b + e$ runs. However the importance of checking fit, blocking, and obtaining an independent estimate of error will differ in different circumstances, and the minimum value of n will thus correspondingly differ. But this minimum design will in any case only be adequate if σ^2 is

below some critical value. When σ^2 is larger designs larger than the minimum design will be needed to obtain estimates of sufficient precision. In this circumstance rather than merely replicate the minimum design, opportunity may be taken to employ a higher order design allowing the fitting of a more elaborate approximating function which can then cover a wider experimental region. Notice that even when σ^2 is small designs for which n is larger than p are not necessarily wasteful. This depends on whether the additional degrees of freedom are genuinely used to achieve the experimenter's current objectives.

Simple Data Patterns

It has sometimes been argued that we may as well choose points randomly to cover the "design region" or employ some algorithm that distributes them evenly even though this does not result in a simple data pattern such as is achieved by factorials and composite response surface designs. In favor of this idea it has been urged that the fitting of a function by least squares to a haphazard set of points is no longer a problem for modern computational devices. This is true, but overlooks an important attribute of designs which form simple patterns. The statistician's task as a member of a scientific team is a dual one, involving inductive criticism and deductive estimation. The latter involves deducing in the light of the data the consequences of given assumptions (estimating the fitted function), and this can certainly be done with haphazard designs. But the former involves the question (a) of what function should be fitted in the first place, and (b) of how to examine

residuals from the fitted function in an attempt to understand deviations from the initial model, in particular in relation to the independent variables, and so to be led to appropriate model modification.

Designs such as factorials and composite response surface designs employ patterns of experimental points that allow many such comparisons to be made, both for the original observations and for the residuals from any fitted function. For example, consider a 3^2 factorial design used to elucidate the effects of temperature and concentration on some response such as yield. Intelligent inductive criticism is greatly enhanced by the possibility of being able to plot the original data and residuals against temperature for each individual level of concentration, and against concentration for each individual level of temperature.

7. CONCLUSION

(i) We must look for good design criteria which measure characteristics of the experimental arrangement in which the scientist might sensibly be interested. Because the importance of various characteristics will differ in different situations, tables of such criteria for particular designs would encourage good judgment to be used in matching the design to the scientific context. Optimum levels of these criteria can be useful as bench marks in judging the efficiencies of a particular design with respect to these various criteria.

(ii) However good designs must in practice be good compromises, and it is doubtful how useful single criterion optimal designs are in locating such compromises. An optimal design is represented by a point in the multi-dimensional space of the coordinates of the design and a series of different

criteria will give a series of such extremal points which can be very differently located. Obviously knowledge of the location of such extrema may tell us almost nothing about the location of good compromises. For this we would need to study the joint behaviour of the criterion functions at levels close to their extremal values. One limited but useful step would be to further investigate which criteria are in accord, (such as G-optimality and robustness to wild observations) and which in conflict (such as variance and bias).

(iii) It is true that the problem of experimental design is full of scientific arbitrariness -- no two investigators would choose the same variables, start their experiments in the same place, change variables over the same regions, and so on -- but science works not by uniqueness but by employing iterative techniques which tend to converge. Clearly we must learn to live with scientific arbitrariness, or else we are in a world of make believe. But we can make the problems worse, not better, by introducing arbitrariness for purely mathematical reasons.

REFERENCES

[1] Atkinson, A.C. (1972). Planning experiments to detect inadequate regression models. Biometrika, 59, 275-293.

[2] Atkinson, A.C. (1973). Multifactor second order designs for cuboidal regions. Biometrika, 60, 15-19.

[3] Atwood, C.L. (1969). Optimal and efficient designs of experiments. Ann. Math. Statist., 40, 1570-1602.

[4] Box, G.E.P. (1954). The exploration and exploitation of response surfaces: Some general considerations and examples. Biometrics, 10, 16-60.

[5] Box, G.E.P. (1968). Response surfaces. Article under 'Experimental Design' in The International Encyclopedia of the Social Sciences, Ed. D.L. Sills, 254-259. New York: MacMillan and Free Press.

[6] Box, G.E.P. and Draper, N.R. (1959). A basis for the selection of a response surface design. J. Am. Statist. Assoc., 54, 622-654.

[7] Box, G.E.P. and Draper, N.R. (1963). The choice of a second order rotatable design. Biometrika, 50, 335-352.

[8] Box, G.E.P. and Draper, N.R. (1975). Robust designs. Biometrika, 62, 347-352.

[9] Box, G.E.P. and Draper, N.R. (1982). Measures of lack of fit for response surface designs and predictor variable transformations. To appear in Technometrics.

[10] Box, G.E.P., Draper, N.R., and Hunter, J.S. Empirical Model Building with Response Surfaces. Book in preparation.

[11] Box, G.E.P. and Hunter, J.S. (1957). Multi-factor experimental designs for exploring response surfaces. Ann. Math. Statist., 28, 195-241.

[12] Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). Statistics for Experimenters. New York: John Wiley & Sons, Inc.

[13] Box, G.E.P. and Lucas, H.L. (1959). Design of Experiments in nonlinear situations. Biometrika, 46, 77-90.

[14] Box, G.E.P. and Wilson, K.B. (1951). On the experimental attainment of optimum conditions. J. R. Statist. Soc., Ser. B, 13, 1-45.

[15] Box, G.E.P. and Youle, P.V. (1955). The exploration and exploitation of response surfaces: An example of the link between the fitted surface and the basic mechanism of the system. Biometrics, 11, 287-323.

[32] Kiefer, J. (1958). On the nonrandomized optimality and the randomized nonoptimality of symmetrical designs. Ann. Math. Statist., 29, 675-699.

[33] Kiefer, J. (1959). Optimum experimental designs. J. R. Statist. Soc., Ser. B, 21, 272-319.

[34] Kiefer, J. (1975). Optimal design: variation in structure and performance under change of criterion. Biometrika, 62, 277-288.

[35] Kiefer, J. and Studden, W.J. (1976). Optimal designs for large degree polynomial regression. Ann. Statist., 4, 1113-1123.

[36] Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. Ann. Math. Statist., 30, 271-294.

[37] Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canadian J. Math., 12, 363-366.

[38] Lucas, J.M. (1976). Which response surface design is best: a performance comparison of several types of quadratic response surface designs in symmetric regions. Technometrics, 18, 411-417.

[39] Mead, R. and Pike, D.J. (1975). A review of response surface methodology from a biometric viewpoint. Biometrics, 31, 803-851.

[40] Nalimov, V.V., Golikovo, T.I., and Mikeshina, N.G. (1970). On practical use of the concept of D-optimality. Technometrics, 12, 799-812.

[41] Pinchbeck, P. H. (1957). The kinetic implications of an empirically fitted yield surface for the vapour phase oxidation of naphthalene to phthalic anhydride. Chem. Engr. Science, 6, 105-111.

[42] Silvey, S.D. and Titterington, D.M. (1973). A geometrical approach to optimal design theory. Biometrika, 60, 21-32.

[43] Titterington, D.M. (1975). Optimal design: some geometrical aspects of D-optimality. Biometrika, 62, 313-320.

[44] Yates, F. (1937). The design and analysis of factorial experiments. Imperial Bureau of Soil Science.

[45] Yates, F. (1967). A fresh look at the basic principles of the design and analysis of experiments. In Proc. 5th Berkeley Symp. Math. Statist. Prob., IV, 777-790. Berkeley and Los Angeles: University of California Press.

[46] Yates, F. (1970). Experimental Design: Selected Papers of Frank Yates. Darien, Connecticut: Hafner.

[16] Box, J.F. (1978). R.A. Fisher: The Life of a Scientist. New York: John Wiley & Sons, Inc.

[17] Box, M.J. (1968). The occurrence of replications in the optimal designs of experiments to estimate parameters in non-linear models. J. R. Statist. Soc., Ser. B, 30, 290-302.

[18] Cochran, W. G. (1973). Experiments for nonlinear functions. J. Am. Statist. Assoc., 68, 771-781.

[19] DeBaun, R. M. (1956). Block effects in the determination of optimum conditions. Biometrics, 12, 20-22.

[20] Draper, N. R. and Lawrence, W. E. (1967). Sequential designs for spherical weight functions. Technometrics, 9, 517-529.

[21] Fedorov, V.V. (1972). Theory of Optimal Experiments. New York and London: Academic Press.

[22] Fisher, R.A. (1921). Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. J. Agric. Sci., 11, 107-135.

[23] Fisher, R.A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. J. Agric. Sci., 13, 311-320.

[24] Fisher, R.A. (1924). The influence of rainfall on the yield of wheat at Rothamsted. Phil. Trans., B, 213, 89-142.

[25] Fisher, R. A. (1935). The Design of Experiments. Oliver and Boyd.

[26] Franklin, N. L., Pinchbeck, P. H., and Popper, F. (1956). A statistical approach to catalyst development, Part 1. The effect of process variables on the vapour phase oxidation of naphthalene. Trans. Inst. Chem. Engrs., 34, 280-293.

[27] Herzberg, A.M. (1979). Are theoretical designs applicable? Operations Research Verfahren/Methods of Operations Research, 30, 68-76.

[28] Herzberg, A.M. (1981). The robust design of experiments: a review. Technical Summary Report 2218, Math. Research Center, University of Wisconsin at Madison.

[29] Herzberg, A.M. and Andrews, D.F. (1976). Some considerations in the optimal design of experiments in non-optimal situations. J. R. Statist. Soc., Ser. B, 38, 284-289.

[30] Herzberg, A.M. and Cox, D.R. (1969). Recent work on the design of experiments: A bibliography and a review. J. R. Statist. Soc., Ser. A, 132, 29-67.

[31] Hill, W.J. and Hunter, W.G. (1966). A review of response surface methodology: a literature survey. Technometrics, 8, 571-590.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: #2333	2. GOVT ACCESSION NO.: AD-A114550
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Choice of Response Surface Design and Alphabetic Optimality	5. TYPE OF REPORT & PERIOD COVERED: Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): George E. P. Box	8. CONTRACT OR GRANT NUMBER(s): DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Work Unit Number 4 - Statistics & Probability
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709	12. REPORT DATE: February 1982
	13. NUMBER OF PAGES: 48
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

D-optimality, A-optimality, E-optimality, G-optimality, response surface designs, information function, bias, sequential assembly, robustness.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is argued that the specification of problems of experimental design (and in particular, of response surface design) should depend on scientific context. The specification for a widely developed theory of "alphabetic optimality" for response surface applications is analyzed and found to be unduly limiting. Ways in which designs might be chosen to satisfy a set of criteria of greater scientific relevance are suggested. Detailed consideration is given to regions of operability and interest, to the design information function, to sensitivity of criteria to size and shape of the region, and to the effect of bias. Problems

(continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT (continued)

are discussed of checking for lack of fit, sequential assembly, orthogonal blocking, estimation of error, estimation of transformations, robustness to bad values, using minimum numbers of points, and employing simple data patterns.

ME

8